

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 16:01:26 ; Search time 5499 Seconds
(without alignments)
11302.766 Million cell updates/sec

Title: US-10-799-747-12
Perfect score: 1434
Sequence: 1 cattaaactcttttctcg.....aaaaaaaaaaaaaaaaaaaaa 1434

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_man.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1432	99.9	1434	6	BD078422	BD078422 101 human
2	1394.8	97.3	3522	9	HSM804706	AL833393 Homo sapi
3	1373.8	95.8	2020	9	HSM800227	AL049442 Homo sapi
4	1368.4	95.4	3452	9	BC033650	BC033650 Homo sapi
c 5	1359.8	94.8	175081	9	AC021151	AC021151 Homo sapi
c 6	591	91.2	281662	2	AC117374	AC117374 Homo sapi
c 7	555	38.7	281662	2	AC117374	AC117374 Homo sapi
c 8	425.8	29.7	448	6	AX408738	AX408738 Sequence
c 9	356.8	24.9	395	6	AX898296	AX898296 Sequence
10	356.8	24.9	395	6	BD033829	BD033829 Sequence
11	84.8	5.9	184359	10	AC121881	AC121881 Mus muscu
c 12	82	5.7	210911	2	AC115265	AC115265 Rattus no
c 13	77.4	5.4	212827	9	AC009412	AC009412 Homo sapi
c 14	76.6	5.3	72832	2	AC099821	AC099821 Homo sapi
c 15	73.2	5.1	110000	3	AC116984_1	Continuation (2 of
c 16	72.2	5.0	2455	3	AF238313	AF238313 Dictyoste
c 17	70.6	4.9	2192	3	AF298207	AF298207 Dictyoste
c 18	70.6	4.9	192929	2	AC005505	AC005505 Plasmodiu
c 19	70.6	4.9	250713	3	AE014850	AE014850 Plasmodiu
c 20	69.8	4.9	545	6	AX185705	AX185705 Sequence
c 21	69.6	4.9	1357	5	BC054578	BC054578 Danio rer
c 22	69.2	4.8	5325	3	AY160094	AY160094 Dictyoste
c 23	69	4.8	8056	6	AX599046	AX599046 Sequence
c 24	68.8	4.8	110000	3	PFMAL1F2_1	Continuation (2 of
c 25	68.4	4.8	110000	3	AC116984_1	Continuation (2 of
c 26	68.2	4.8	64707	3	AC115607	AC115607 Dictyoste
c 27	68.2	4.8	254436	3	AE014827	AE014827 Plasmodiu
c 28	68	4.7	189218	2	EX897692	EX897692 Danio rer
c 29	67.8	4.7	302156	3	AC116977	AC116977 Dictyoste
c 30	67.6	4.7	125623	3	AC115599	AC115599 Dictyoste
c 31	67.6	4.7	182870	3	AC116960	AC116960 Dictyoste
c 32	67.6	4.7	266544	3	AC116956	AC116956 Dictyoste
c 33	67.4	4.7	125958	3	AC115592	AC115592 Dictyoste
c 34	67	4.7	110000	2	PFMAL8P1_00	AL844507 Plasmodiu
c 35	67	4.7	110000	2	PFMAL8P1_01	Continuation (2 of
c 36	66.6	4.6	1092	3	DDSTATCUL	Y13098 Dictyosteli
c 37	66.6	4.6	254436	3	AE014827	AE014827 Plasmodiu
c 38	66.2	4.6	110000	2	PFMAL7P1_03	Continuation (4 of
c 39	66	4.6	166447	2	BX323824	BX323824 Danio rer
c 40	65.6	4.6	597	6	AX187052	AX187052 Sequence
c 41	65.6	4.6	2067	3	FFA271108	AX271108 Plasmodiu
c 42	65.6	4.6	254733	3	AC117075	AC117075 Dictyoste
c 43	65.4	4.6	110000	2	PFMAL13_19	Continuation (20 o
c 44	65.4	4.6	136240	3	AC117070	AC117070 Dictyoste
c 45	65.2	4.5	41399	3	AC116984_5	Continuation (6 of

ALIGNMENTS

RESULT 1	BD078422	101 human secretory proteins.	1434 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	BD078422					
DEFINITION	BD078422					
ACCESSION	BD078422					
VERSION	BD078422.1	GI:22624025				
KEYWORDS	JP 2001519156-A/11					
SOURCE	Homo sapiens (human)					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1 (bases 1 to 1434)					
AUTHORS	Duan,R.D., Florence,K.A., Rosen,C.A., Ruben,S.M., Greene,J.M.,					
	Young,P., Ferrie,A.M., Yu,G.L., Janat,F.Ni,J., Carter,K.C.,					
	Endress,G.A., Feng,P., Lafleur,D.W. and Shi,Y.					

TITLE	101 human secretory proteins
JOURNAL	Patent: JP 2001519156-A 11 23-OCT-2001;
COMMENT	HUMAN GENOME SCIENCES INC OS Homo sapiens (human) PN JP 2001519156-A/11 PD 23-OCT-2001 PF 01-OCT-1998 JP 2000515006 PR 02-OCT-1997 US 60/060837, 02-OCT-1997 US 60/060862 PR 02-OCT-1997 US 60/060839, 02-OCT-1997 US 60/060866 PR 02-OCT-1997 US 60/060843, 02-OCT-1997 US 60/060836 PR 02-OCT-1997 US 60/060838, 02-OCT-1997 US 60/060874 PR 02-OCT-1997 US 60/060833, 02-OCT-1997 US 60/060884 PR 02-OCT-1997 US 60/060880 PI ROXANNE D DUAN, KIMBERLY A FLORENCE, CRAIG A ROSEN, STEVEN M PI RUBEN, PI JOHN M GREENE, PAUL YOUNG, ANN M FERRIE, GUO LIANG YU, FOUAD JANAT, PI JIAN NI, PI KENNETH C CARTER, GREGORY A ENDRESS, PING FENG, DAVID W LAFLEUR, PI YANGGU SHI PC C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P25/00, A61P35/00, PC A61P37/00, PC A61P43/00, C07K14/47, C07K16/00, C12N5/10, C12P21/02, C12Q1/68, PC G01N33/53, PC G01N33/53, G01N33/566, C12N15/00, A61K37/02, C12N5/00 CC 101 human secretory proteins FH Key Location/Qualifiers FT source 1. .1434 /organism="Homo sapiens (human)". FT 1. .1434 Location/Qualifiers FT 1. .1434 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
FEATURES	source
ORIGIN	
Query Match	99.9%; Score 1432; DB 6; Length 1434;
Best Local Similarity	100.0%; Pred. No. 2.3e-227;
Matches 1434; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 CATTAACTCTTTTATCGGAATAGTATGATATTTCAATGTCACCTCCATTCATGTTGA 60
Db	1 CATTAACTCTTTTATCGGAATAGTATGATATTTCAATGTCACCTCCATTCATGTTGA 60
QY	61 TTTGGAGTCACAGTATTTTGTGTAAGCAGAGATTTAAATTTATTTGTAAGTCAGTGC 120
Db	61 TTTGGAGTCACAGTATTTTGTGTAAGCAGAGATTTAAATTTATTTGTAAGTCAGTGC 120
QY	121 AAAATTATGAATAGGATATCTAATAATACAAAGTAATAACAAAGTCACAAAGCAGTGT 180
Db	121 AAAATTATGAATAGGATATCTAATAATACAAAGTAATAACAAAGTCACAAAGCAGTGT 180
QY	181 CTAATAAATAATCTGGGTTCTTAAATAATTTTAAATTTATCTTTGAAATAGTTTCT 240
Db	181 CTAATAAATAATCTGGGTTCTTAAATAATTTTAAATTTATCTTTGAAATAGTTTCT 240
QY	241 TAGAATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 300
Db	241 TAGAATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 300
QY	301 CAAATTGCTATTAAATGCMAGAGTGTTAATATACAGAAATTTATCAGGCATTACCAAGTC 360
Db	301 CAAATTGCTATTAAATGCMAGAGTGTTAATATACAGAAATTTATCAGGCATTACCAAGTC 360
QY	361 TAGGCATATAGGAATSCAGACCTCAGAAATGGTTTCAATGTAGTATGATGCTTGTA 420
Db	361 TAGGCATATAGGAATSCAGACCTCAGAAATGGTTTCAATGTAGTATGATGCTTGTA 420
QY	421 AGGTAGGGAGCTTATTTCAGACATAGTAGTATTTCTCTAAATGCTGTSTCAATTTGCTGG 480
Db	421 AGGTAGGGAGCTTATTTCAGACATAGTAGTATTTCTCTAAATGCTGTSTCAATTTGCTGG 480
QY	481 CCTTTGGCTACCTGTACTTCCSATTATGGAGCCCAATTCAGTCTTCAGTCTTCTCTCT 540

Db	481 CCTTTGGCTACCTGTACTTCCSATTATGGAGCCCAATTCAGTCTTTCAGTCTTCTCTCT 540
QY	541 GGCACACTTATGCTCTGAAATCATGAGCGAGGCTGATTTCAATTTGGTGTAGTAAAG 600
Db	541 GGCACACTTATGCTCTGAAATCATGAGCGAGGCTGATTTCAATTTGGTGTAGTAAAG 600
QY	601 GCAGTATGTTTGTCTGACATTAAGATGTAGTTATAGATAGGTTTACGCTTTAAGTGTAT 660
Db	601 GCAGTATGTTTGTCTGACATTAAGATGTAGTTATAGATAGGTTTACGCTTTAAGTGTAT 660
QY	661 GTTTTTATATCTTTTAAATAAGAAATAAACCCTTTTAAAGCTATTCCACCTCTCTCCCCAGC 720
Db	661 GTTTTTATATCTTTTAAATAAGAAATAAACCCTTTTAAAGCTATTCCACCTCTCTCCCCAGC 720
QY	721 CTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAACCTTCAC 780
Db	721 CTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAACCTTCAC 780
QY	781 GCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTTAATAC 840
Db	781 GCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTTAATAC 840
QY	841 TTACTTCTATTTCTGAAATCAGTTTGAACTGTTGCTTTATGTTCAAGAGTTTAAAGAAC 900
Db	841 TTACTTCTATTTCTGAAATCAGTTTGAACTGTTGCTTTATGTTCAAGAGTTTAAAGAAC 900
QY	901 CTCMGTAATTCATTTTAAATCTGCTATTCGAGAAGCATTTGAATGAATTTTAAACA 960
Db	901 CTCMGTAATTCATTTTAAATCTGCTATTCGAGAAGCATTTGAATGAATTTTAAACA 960
QY	961 AGAAGCTCATCTGAGTCTTTTGGCTGACCTCTATGAGCCCATAGGGTCTCTGCTTA 1020
Db	961 AGAAGCTCATCTGAGTCTTTTGGCTGACCTCTATGAGCCCATAGGGTCTCTGCTTA 1020
QY	1021 GCATTAACAAATAAGGTTTATAGTAAAGCCCAATGTAATTTTTCATGCTGAGG 1080
Db	1021 GCATTAACAAATAAGGTTTATAGTAAAGCCCAATGTAATTTTTCATGCTGAGG 1080
QY	1081 GCTTTAAATTTGCTCTCTTTTCAATTTTATTTATTTATTTATTTGTTGTAAGT 1140
Db	1081 GCTTTAAATTTGCTCTCTTTTCAATTTTATTTATTTATTTATTTGTTGTAAGT 1140
QY	1141 CTTTTTGGGAGATAATATATGTTTAAATTTAGTTTGGGGGGAATAATTTGTCGAAGA 1200
Db	1141 CTTTTTGGGAGATAATATATGTTTAAATTTAGTTTGGGGGGAATAATTTGTCGAAGA 1200
QY	1201 GGATAATTTAAATTTACGTGCTTCTGTTATTCAGAAATAAGAGAGAGACTACGCTGCATA 1260
Db	1201 GGATAATTTAAATTTACGTGCTTCTGTTATTCAGAAATAAGAGAGAGACTACGCTGCATA 1260
QY	1261 TTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTCTTAAGATTTTCAAAAGGAATAT 1320
Db	1261 TTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTCTTAAGATTTTCAAAAGGAATAT 1320
QY	1321 GTGTAAATTCAGAAATCATCAACCACTGCTCTAACTTGGTAAACAACTGTTCTTAATAA 1380
Db	1321 GTGTAAATTCAGAAATCATCAACCACTGCTCTAACTTGGTAAACAACTGTTCTTAATAA 1380
QY	1381 AGTATTTAAATGATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1434
Db	1381 AGTATTTAAATGATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1434

RESULT 2
 HSM804706
 LOCUS
 DEFINITION Homo sapiens mRNA; cDNA DKFp762K109 (from clone DKFp762K109).
 ACCESSION AL833393
 VERSION AL833393.1
 KEYWORDS GI:21734029
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 3522)

Ottenwaelder, B., Obermaier, B., Mewes, H.W., Weil, B. and Wiemann, S. Direct Submission

Submitted (09-JUL-2002) 1, D-85764 Neuberg, GERMANY

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by MediGenomix (Martinried/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp762K109) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

FEATURES

Location/Qualifiers

1..3522

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="RZPD:DKFZp762K109"

/db_xref="taxon:9606"

/clone="DKFZp762K109"

/tissue_type="melanoma (MeWo cell line)"

/clone_lib="762 (synonym: hmel2). Vector pSport1; host DH10B; sites NotI + SalI"

/dev_stage="adult"

polyA_signal 3420..3425

polyA_site 3440

ORIGIN

Query Match 97.3%; Score 1394.8; DB 9; Length 3522;

Best Local Similarity 99.3%; Pred. No. 2.8e-221;

Matches 1426; Conservative 5; Mismatches 2; Indels 3; Gaps 3;

QY 1 CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTCACCTCCATTCATGTGA 60

DB 2044 CATTAACCTCTTTTATCGGGAATAGTATGATATTTTCAATGTCACCTCCATTCATGTGA 2103

QY 61 TTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTTAAATTTATATTTGAAGTCAGTGC 120

DB 2104 TTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTTAAATTTATATTTGAAGTCAGTGC 2163

QY 121 AAAATTATGAATAGGATATCTAATAAATAAAGTAATTAACAAAGTCAAAGTCAGTGT 180

DB 2164 AAAATTATGAATAGGATATCTAATAAATAAAGTAATTAACAAAGTCAAAGTCAGTGT 2223

QY 181 CTAAATAAAATTCCTGGTTCCTTAAATAATTTTAAATTTATC-TTGAATAGTTTTC 239

DB 2224 CTAAATAAAATTCCTGGTTCCTTAAATAATTTTAAATTTATC-TTGAATAGTTTTC 2283

QY 240 TTAGATTAAATCTCAGGATATGAAAGTCATTAATTAAGTGTGAGTAAAGTTAGTATCATTA 299

DB 2284 TTAGATTAAATCTCAGGATATGAAAGTCATTAATTAAGTGTGAGTAAAGTTAGTATCATTA 2343

QY 300 ACAAATGCTCTAATAATGACAG-GTGGTAATATACAGAAATTAATACAGCAATACCAAG 358

DB 2344 ACAAATGCTCTAATAATGACAGCGTGGTAATATACAGAAATTAATACAGCAATACCAAG 2403

QY 359 TCTAGGCACATATAGGAATGACGACCTCAGATGTTTCAATGTAGTAGTTCATGCTTG 418

DB 2404 TCTAGGCACATATAGGAATGACGACCTCAGATGTTTCAATGTAGTAGTTCATGCTTG 2463

QY 419 TAAGGTAGGGAGCTTATTCAGACATAGTAGTAGTTTCTCTAATGCTGTSTCAATTTGCT 478

DB 2464 TAAGGTAGGGAGCTTATTCAGACATAGTAGTAGTTTCTCTAATGCTGTCTCAATTTGCT 2523

QY 479 GGCCTTTGGCTACTGCTACTTCSCATTATGGCAGCCATTCAGTCTTGATTTCTTCT 538

DB 2524 GGCCTTTGGCTACTGCTACTTCSCATTATGGCAGCCATTCAGTCTTGATTTCTTCT 2583

QY 539 CTGGACACCTTATGCTCTGAAATCATGAGCGAGCTGATTCAATTTGGTGAATTTGGGTAGA 598

DB 2584 CTGGACACCTTATGCTCTGAAATCATGAGCGAGCTGATTCAATTTGGTGAATTTGGGTAGA 2643

QY 599 AAGCAGTATCTTTTGTGTCACATTAAGATGTAGGTTATAGATAGGTTTACGCTTTAAGTGT 658

DB 2644 AAGCAGTATCTTTTGTGTCACATTAAGATGTAGGTTATAGATAGGTTTACGCTTTAAGTGT 2703

QY 659 ATGTTTTTATACCTTTTAAATAAATAAATACCTTTTAAAGCTATTCACCTCTCCCTCCCA 718

DB 2704 ATGTTTTTATACCTTTTAAATAAATAAATACCTTTTAAAGCTATTCACCTCTCCCTCCCA 2763

QY 719 GCCTATCTCAAACTGTTGGGAATATATGGAGAGATCTTGAAGAGTAAATAAATCAACCTTCA 778

DB 2764 GCCTATCTCAAACTGTTGGGAATATATGGAGAGATCTTGAAGAGTAAATAAATCAACCTTCA 2823

QY 779 CTGCTCCACTCCAGGTGAATCCGCCACTCCCTGACCTGACCTAGTAGAATTTGTATTTAAT 838

DB 2824 CTGCTCCACTCCAGGTGAATCCGCCACTCCCTGACCTGACCTAGTAGAATTTGTATTTAAT 2883

QY 839 ACTTACCTTCTATTTCTGAATCAGTTGTGAACCTGTTGCCTTATGTTTCAGAGTTTAAAGA 898

DB 2884 ACTTACCTTCTATTTCTGAATCAGTTGTGAACCTGTTGCCTTATGTTTCAGAGTTTAAAGA 2943

QY 899 ACCTCMGTGAATTCATTTTAAATAAATCTGCTATTTCTGAGAGCATTGAATGAATTTCTTAA 958

DB 2944 ACCTCMGTGAATTCATTTTAAATAAATCTGCTATTTCTGAGAGCATTGAATGAATTTCTTAA 3003

QY 959 CAAGAAGACTCATCTGTAGTCTGTTGCTGATCTCTATGAGCCCTCAATAAGGTTCTGTGCT 1018

DB 3004 CAAGAAGACTCATCTGTAGTCTGTTGCTGATCTCTATGAG-CCATAAAGGTTCTGTGCT 3062

QY 1019 TAGCATTAACAAAATAAGTTTATAGTAAAGCAATGTTAATTTTTTTTTTTCATGGA 1078

DB 3063 TAGCATTAACAAAATAAGTTTATAGTAAAGCAATGTTAATTTTTTTTTTTCATGGA 3122

QY 1079 GGGCTTTAAATTTGCTCTTTTTCATATTTTATTCATATTCATATTTATGTTTGTGTAAC 1138

DB 3123 GGGCTTTAAATTTGCTCTTTTTCATATTTTATTCATATTCATATTTATGTTTGTGTAAC 3182

QY 1139 TGCTTTTTAGGGAGATAAATATATATGTTTAAATTTAGTTTTGGGGGAATAATTTGTGCAAA 1198

DB 3183 TGCTTTTTAGGGAGATAAATATATATGTTTAAATTTAGTTTTGGGGGAATAATTTGTGCAAA 3242

QY 1199 GAGGATAATTTAATTTACGTGCTTCTGTTTATTCAGATAAAGAGAGAGACCTACGCTGCA 1258

DB 3243 GAGGATAATTTAATTTACGTGCTTCTGTTTATTCAGATAAAGAGAGAGACCTACGCTGCA 3302

QY 1259 TATTCAAGAGTCTGACTTAAACATTCGTGAAACATTTTTTCTAAGATTTTCAAAAGCAAT 1318

DB 3303 TATTCAAGAGTCTGACTTAAACATTCGTGAAACATTTTTTCTAAGATTTTCAAAAGCAAT 3362

QY 1319 ATGTGAAATTTGAGAAATCATACCACTGCTTAACTTGGTAAACAACTGTTCTTAAAT 1378

DB 3363 ATGTGAAATTTGAGAAATCATACCACTGCTTAACTTGGTAAACAACTGTTCTTAAAT 3422

QY 1379 AAAGTATTTAATGATTTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAA 1434

DB 3423 AAAGTATTTAATGATTTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAA 3478

RESULT 3

HSM800227

LOCUS

DEFINITION

ACCESSION

AL049442

VERSION

AL049442.1

GI:4500222

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 2020)

Wambutt, R., Heubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S. Direct Submission

Submitted (10-MAR-1999) MIPS, Am Klopferspitz 18a, D-82152

Martinsried, GERMANY
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOMA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
 This clone (DKFZ586N1720) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.
 Location/Qualifiers
 1. .2020
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="RZPD:DKFZ586N1720"
 /db_xref="taxon:9606"
 /clone="DKFZ586N1720"
 /tissue_type="uterus"
 /clone_lib="586 (synonym: hutel). Vector pSport1; host DH10B; sites NotI + SalI/MluI"
 /dev_stage="adult"
 1983. .1988
 2003
 polyA_signal
 polyA_site
 ORIGIN
 Query Match 95.8%; Score 1373.8; DB 9; Length 2020;
 Best Local Similarity 95.3%; Pred. No. 9.2e-218; Mismatches 5; Conservative 2; Indels 3; Gaps 3;
 Matches 1405;

Qy 1 CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTGCACATCCCATTCATGTTGA 60
 Db 607 CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTGCACATCCCATTCATGTTGA 666
 Qy 61 TTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTTAAATTTATATGTAAGTCAGTGC 120
 Db 667 TTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTTAAATTTATATGTAAGTCAGTGC 726
 Qy 121 AAAATTATGATAGATATCTAATAATACAAAGTAAATACAAAGTCAAAAGCAGAGTT 180
 Db 727 AAAATTATGATAGATATCTAATAATACAAAGTAAATACAAAGTCAAAAGCAGAGTT 786
 Qy 181 CTAATAAAAATCTCGGTTCTTTAAATAATTTTAAATTTATC-TTGAATAAGTTTTC 239
 Db 787 CTAATAAAAATCTCGGTTCTTTAAATAATTTTAAATTTATC-TTGAATAAGTTTTC 846
 Qy 240 TTAGATTAACTCAGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA 299
 Db 847 TTAGATTAACTCAGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA 906
 Qy 300 ACAATTGTCTATTAAATGCAMGA-GTGGTAAATACAGAAATTTATCAGGCATTACCAAG 358
 Db 907 ACAATTGTCTATTAAATGCAMGACGTGGTAAATACAGAAATTTATCAGGCATTACCAAG 966
 Qy 359 TCTAGGCACATATAGGAATGCAGCATCAGAAATGGTTTCAAGTAGTATGATGCTTG 418
 Db 967 TCTAGGCACATATAGGAATGCAGCATCAGAAATGGTTTCAAGTAGTATGATGCTTG 1026
 Qy 419 TAAGGTAGGGAGCTTATTCAGATAGTATGTTTCTCTAATGCTGTCTCAATGCT 478
 Db 1027 TAAGGTAGGGAGCTTATTCAGATAGTATGTTTCTCTAATGCTGTCTCAATGCT 1086
 Qy 479 GGCCTTTGGTACCTGTACTTCCSCATTATGGCAGCCCAATTCAGTCTTGAGTTTCTTCT 538
 Db 1087 GGCCTTTGGTACCTGTACTTCCCAATATGGCAGCCCAATTCAGTCTTGAGTTTCTTCT 1146
 Qy 539 CTGACACCTTATGCTCTGAATCATGAGCAGGCTGATTCATTAAGTGTGATTTGGGTAGA 598
 Db 1147 CTGACACCTTATGCTCTGAATCATGAGCAGGCTGATTCATTAAGTGTGATTTGGGTAGA 1206
 Qy 599 AAGCAGTATGTTTGTGACATTAAGATGTAGGTTATAGATAGTGTGTAGCTTTAAGTGT 658
 Db 1207 AAGCAGTATGTTTGTGACATTAAGATGTAGGTTATAGATAGTGTGTAGCTTTAAGTGT 1266

Qy 659 ATGTTTTTATCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCCACCTCTCCCTCCCA 718
 Db 1267 ATGTTTTTATCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCCACCTCTCCCTCCCA 1326
 Qy 719 GCCTATCTCAAACTGGTGGATATATGAGAGAGATCTTGAAGAAGTAAATAAACCCTTCA 778
 Db 1327 GCCTATCTCAAACTGGTGGATATATGAGAGAGATCTTGAAGAAGTAAATAAACCCTTCA 1386
 Qy 779 CTGCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAAATTAAT 838
 Db 1387 CTGCTCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAAATTAAT 1446
 Qy 839 ACTTACCTCTTATTTCTGAAATCAGTTGTGAACCTGTTCCTATGTTTCAGARGTTTGA 898
 Db 1447 ACTTACCTCTTATTTCTGAAATCAGTTGTGAACCTGTTCCTATGTTTCAGARGTTTGA 1506
 Qy 899 ACCTCMGTGAATTCATTTTAAATCTGCTATTTCTGAGAAGCATTCGAATGAATCTTAA 958
 Db 1507 ACCTCAGTGAATTCATTTTAAATCTGCTATTTCTGAGAAGCATTCGAATGAATCTTAA 1566
 Qy 959 CAAGAAGACTCATCTGTAGCTGTTTGTGACTCTCTATGAGCCCATAGGGTCTGTGCT 1018
 Db 1567 CAAGAAGACTCATCTGTAGCTGTTTGTGACTCTCTATGAG-CCCATAGGGTCTGTGCT 1625
 Qy 1019 TAGCATTAACAAAATAAGGTTTATAGCTAAAGCCAAATGATTAATTTTTTTTTCATGGA 1078
 Db 1626 TAGCATTAACAAAATAAGGTTTATAGCTAAAGCCAAATGATTAATTTTTTTTTCATGGA 1685
 Qy 1079 GGGCTTTAAATTTGTGCTCTTTTTCATATTTTATTCATATTCATATTCATATTCATAT 1138
 Db 1686 GGGCTTTAAATTTGTGCTCTTTTTCATATTTTATTCATATTCATATTCATATTCATAT 1745
 Qy 1139 TGCTTTTACGGAGATTAATATATGTTTAAATAGTTTGGGGGGAATTAATTTGTGCAA 1198
 Db 1746 TGCTTTTACGGAGATTAATATATGTTTAAATAGTTTGGGGGGAATTAATTTGTGCAA 1805
 Qy 1199 GAGGATAATTTAATTTACGTGCTTCTGTTATTCAGAAATAAGAGAGAGACTACGCTGCA 1258
 Db 1806 GAGGATAATTTAATTTACGTGCTTCTGTTATTCAGAAATAAGAGAGAGACTACGCTGCA 1865
 Qy 1259 TATTCAGAGTTGTACCTTAACATTTGGTGAAACATTTTTCCTAAGATTTTCAAAGGAAT 1318
 Db 1866 TATTCAGAGTTGTACCTTAACATTTGGTGAAACATTTTTCCTAAGATTTTCAAAGGAAT 1925
 Qy 1319 ATGTGTAATTTGAGAAATCATACCACTGCTTAACCTTGTGTAACCAACTGTTCTTAAT 1378
 Db 1926 ATGTGTAATTTGAGAAATCATACCACTGCTTAACCTTGTGTAACCAACTGTTCTTAAT 1985
 Qy 1379 AAGTATTTAATGATTTTAAAAAATAAAAAA 1413
 Db 1986 AAGTATTTAATGATTTTAAAAAATAAAAAA 2020

RESULT 4
 BC033650 3452 bp mRNA linear PRI 06-OCT-2003
 LOCUS Homo sapiens hypothetical protein FLJ14431, mRNA (cDNA clone MGC:44889 IMAGE:5574637), complete cds.
 ACCESSION BC033650
 VERSION BC033650.1 GI:21707066
 KEYWORDS MGC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 3452)
 AUTHORS Klausner,R.D., Collins,F.S., Wagner,L., Shennan,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,I.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,

Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, K.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalhus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.,
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
22388257
12477932
2 (bases 1 to 3452)
Strausberg, R.
Direct Submission
Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapsb-xemail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hgrl.nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghghi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 69 Row: n Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 14249445.

FEATURES

Location/Qualifiers
1..3452
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:44889 IMAGE:5574637"
/tissue_type="Duodenum, adenocarcinoma"
/clone_lib="NIH_MGC_88"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
1..3452
/gene="FLJ14431"
/db_xref="LocusID:84869"
159..872
/codon_start=1
/product="FLJ14431 protein"
/protein_id="AAH33650.1"
/db_xref="GI:21707067"
/db_xref="LocusID:84869"
/translations="MDKVCVAVFGSGRIGRAVALMARKGYRLAVIARNLEGAKAAG DLGGDLAFSCDVAKHEDVQNTPEELEKHLGRVFNVAAGINRDLVLRKTKEDMVS QLHTNLIGSMILTCKAAMRTMIOQGGSI VNVGSI VGLKNGSGSVY SASKGGLVFSR ALAKAVARKKIRVNVVAPGFVHTDMTKDLKEHLKKNIPILGRFGEIEVAHVFFLE SPVITGHVLVVDGGLQLIL"
159..857
misc_feature

/notes="adh_short; Region: short chain dehydrogenase. This family contains a wide variety of dehydrogenases"
/db_xref="CDD:pfam00106"
ORIGIN
Query Match 95.4%; Score 1368.4; DB 9; Length 3452;
Best Local Similarity 99.3%; Pred. No. 6.5e-217;
Matches 1410; Conservative 5; Mismatches 1; Indels 4; Gaps 4;
QY 1 CATTAAACTCTTTTATCGGGATAGTATGATATTTTCAATGTCACCTCCATTCATGTTGA 60
Db 2034 CATTAACTCTTTTATCGGGATAGTATGATATTTTCAATGTCACCTCCATTCATGTTGA 2093
QY 61 TTTGGAGCTGACAGTTATTTTGTGTGAAGCAGAGATTAAATTTTATATGAAAGTCAGTGC 120
Db 2094 TTTGGAGCTGACAGTTATTTTGTGTGAAGCAGAGATTAAATTTTATATGAAAGTCAGTGC 2153
QY 121 AAAATTATGAATAGGATATCTAATAATAACAAAGTAATAACAAAGTCAAAAGTCAGAGTGT 180
Db 2154 AAAATTATGAATAGGATATCTAATAATAACAAAGTAATAACAAAGTCAAAAGTCAGAGTGT 2213
QY 181 CTAATAAATAAATTCCTGGTTCCTTAAAAATTTATTTAAAAATTTATC-TTGAATAAGTTTTC 239
Db 2214 CTAATAAATAAATTCCTGGTTCCTTAAAAATTTATTTAAAAATTTATC-TTGAATAAGTTTTC 2273
QY 240 TTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA 299
Db 2274 TTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTA 2333
QY 300 ACAAATGTCATTAATGCMGA-GTGGTAATATACAGAAATTTATCAGGCATTTACCAAG 358
Db 2334 ACAAATGTCATTAATGCMGACGTGGTAATATACAGAAATTTATCAGGCATTTACCAAG 2393
QY 359 TCTAGGCACATATAGAAATGACGACTCAGAAATGTTTCAATGTAAGTGTGAGTGTG 418
Db 2394 TCTAGGCACATATAGAAATGACGACTCAGAAATGTTTCAATGTAAGTGTGAGTGTG 2453
QY 419 TAAAGTAGGGAGCTTATTACAGACATAGTAGATAGTTTCTTAATGCTGTSTCAATGCT 478
Db 2454 TAAAGTAGGGAGCTTATTACAGACATAGTAGATAGTTTCTTAATGCTGTCTCAATGCT 2513
QY 479 GGCTTTGGTACTGCTTCCSCATTATGGCAGGCCATTCAGTCTTGAGTTTCTTCT 538
Db 2514 GGCTTTGGTACTGCTTCCSCATTATGGCAGGCCATTCAGTCTTGAGTTTCTTCT 2573
QY 539 CTGCACACCTTATGCTCTGAAATCATGAGCGAGCTGATTCAATTTGGTGAATTTGGGTAGA 598
Db 2574 CTGCACACCTTATGCTCTGAAATCATGAGCGAGCTGATTCAATTTGGTGAATTTGGGTAGA 2633
QY 599 AAGCAGTATGTTTGTGTCATTAAGATGTAGGTATAGATAGTTTAGCCTTTAAGTGT 658
Db 2634 AAGCAGTATGTTTGTGTCATTAAGATGTAGGTATAGATAGTTTAGCCTTTAAGTGT 2693
QY 659 ATGTTTTTATCTTAAATAAGAAATATAACCTTTTAAGCTATTCCACCTCTCTCCCCCA 718
Db 2694 ATGTTTTTATCTTAAATAAGAAATATAACCTTTTAAGCTATTCCACCTCTCTCCCCCA 2753
QY 719 GCCTATCTCAAACTGCTGGAAATATATGGAGAGATCTTGAAGAGTAATAAAACCTTCA 778
Db 2754 GCCTATCTCAAACTGCTGGAAATATATGGAGAGATCTTGAAGAGTAATAAAACCTTCA 2813
QY 779 CTGCTCCACTCCAGGTGAATCCGCCCACTCCCACTGACCTAGTATGTAATTTAAT 838
Db 2814 CTGCTCCACTCCAGGTGAATCCGCCCACTCCCACTGACCTAGTATGTAATTTAAT 2873
QY 839 ACTTACTTCTATTCTGAAATCAGTTGTGAACCTGTGCTTATGTTTCAGAGTTTAA 898
Db 2874 ACTTACTTCTATTCTGAAATCAGTTGTGAACCTGTGCTTATGTTTCAGAGTTTAA 2933
QY 899 ACCTCMGTGAATTCATTTTTTAAAACTGCTATTCTTGAGAAGCATTGAATGAATTTCTTA 958
Db 2934 ACCTCAGTAATTCATTTTTTAAAACTGCTATTCTTGAGAAGCATTGAATGAATTTCTTA 2993

```

QY 959 CAAGAAGACTCACTGTAGCTGTTTCTGCTGACTCCTATGAGCCCATAGGGTTCCTGTGCT 1018
Db |||||||
QY 2994 CAAGAAGACTCACTGTAGCTGTTTCTGCTGACTCCTATGAG-CCCATAGGGTTCCTGTGCT 3052
Db |||||||
QY 1019 TAGCATTAACAAAAATAAGGTTTATAGGTAAAGCCCAATGTAATATTTTTCGTCATGGA 1078
Db |||||||
QY 3053 TAGCATTAACAAAAATAAGGTTTATAGGTAAAGCCCAATGTAATATTTTTCGTCATGGA 3112
Db |||||||
QY 1079 GGGCTTTAAATTTGCTGCTCTTTTTCATATTTTATTCATATTCATATTTGTTGTAAC 1138
Db |||||||
QY 3113 GGGCTTTAAATTTGCTGCTCTTTTTCATATTTTATTCATATTCATATTTGTTGTAAC 3172
Db |||||||
QY 1139 TGCTTTTTCAGGAGATAATATATATGTTATATATTTAGTTTTCGGGGGGAATATTCGTCGAAA 1198
Db |||||||
QY 3173 TGCTTTTTCAGGAGATAATATATATGTTATATATTTAGTTTTCGGGGGGAATATTCGTCGAAA 3232
Db |||||||
QY 1199 GAGGATAATTTAAATTTGCTGCTCTCTGTTATTCAGATAAAGAGAGAGACTACGCTGCA 1258
Db |||||||
QY 3233 GAGGATAATTTAAATTTGCTGCTCTCTGTTATTCAGATAAAGAGAGAGACTACGCTGCA 3292
Db |||||||
QY 1259 TATTCAGAGTTGTACCTTAACATTTGTTGAACA-TTTTCTTAAGATTTTCAAAAGGAA 1317
Db |||||||
QY 3293 TATTCAGAGTTGTACCTTAACATTTGTTGAACA-TTTTCTTAAGATTTTCAAAAGGAA 3352
Db |||||||
QY 1318 TATGCTAAATTTGAGAAATCATAACCTGCTCACTGTTGTAACAAACCTGTTCTTAAA 1377
Db |||||||
QY 3353 TATGCTAAATTTGAGAAATCATAACCTGCTCACTGTTGTAACAAACCTGTTCTTAAA 3412
Db |||||||
QY 1378 TAAAGTATTTAATGATTTTAAAAA-AAAAAAAAAAAA 1417
Db |||||||
QY 3413 TAAAGTATTTAATGATTTTAAAAA-AAAAAAAAAAAA 3452
Db |||||||

RESULT 5
AC021151/c 175081 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-483A20 from 4, complete sequence.
DEFINITION AC021151
ACCESSION AC021151
VERSION AC021151.8 GI:15145598
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
93063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 175081)
AUTHORS Kyung, K. and Abbott, A.
TITLE The sequence of Homo sapiens BAC clone RP11-483A20
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 175081)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 175081)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 175081)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Aug 9, 2001 this sequence version replaced gi:1387722.
----- Genome Center

```

Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.edu
----- Summary Statistics
Center project name: H_NH0483A20

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-36G9. Actual start of this clone is at base position 1 of RP11-483A20; actual end is at base position 175081 of RP11-483A20.

FEATURES

source	1..175081
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-483A20"
	/clone_lib="RP11-11"
repeat_region	1..3267
	/rpt_family="L1"
repeat_region	3270..3453
	/rpt_family="Alu"
repeat_region	3434..3464
	/rpt_family="(A)n"
repeat_region	3455..4161
	/rpt_family="L1"
misc_feature	7949..7956
	/note="match to EST AW006057 (NID:g5854835) wz8ld04.x1"
	/10197..10489
repeat_region	10197..10489
	/rpt_family="L1"
repeat_region	10480..10501
	/rpt_family="AT-rich"
repeat_region	10495..10665
	/rpt_family="L1"
repeat_region	10731..11051
	/rpt_family="L1"
repeat_region	11052..11576
	/rpt_family="L1"
repeat_region	11564..11598
	/rpt_family="(A)n"
repeat_region	11642..12324

```
/rpt_family="L1"
11769..11794
/rpt_family="AT_rich"
11939..11962
/rpt_family="AT_rich"
12335..12412
/rpt_family="L1"
12392..12621
/rpt_family="L1"
12788..13096
/rpt_family="Alu"
13069..13102
/rpt_family="(CAAAA)n"
13325..13667
/rpt_family="Alu"
13444..13484
/rpt_family="(TAAA)n"
13640..13686
/rpt_family="(GAAA)n"
14249..14422
/rpt_family="MER1_type"
14980..15182
/rpt_family="L1"
15260..15482
/rpt_family="L1"
15483..15985
/rpt_family="MER1_type"
15986..16105
/rpt_family="L1"
16117..16251
/rpt_family="L1"
16288..16356
/rpt_family="L1"
16373..16545
/rpt_family="L1"
16550..18773
/rpt_family="L1"
18770..20218
/rpt_family="L1"
19228..19346
/rpt_family="T-rich"
20219..20514
/rpt_family="Alu"
20515..22484
/rpt_family="L1"
22584..22882
/rpt_family="Alu"
22748..22768
/rpt_family="AT_rich"
22883..23460
/rpt_family="L1"
23461..23747
/rpt_family="MaLR"
23749..24514
/rpt_family="L1"
24456
/notes="match to EST AW902102 (NID:g8066307)"
24488..24797
/notes="match to EST AW902102 (NID:g8066307)"
24576..24625
/rpt_family="Mariner"
25484..25787
/rpt_family="Alu"
25758..25787
/rpt_family="(A)n"
25796..26014
/rpt_family="MIR"
26657..26889
/rpt_family="L1"
26890..27205
/rpt_family="MaLR"
27206..27439
/rpt_family="Alu"
27440..27490
/rpt_family="MaLR"
27491..28717
/rpt_family="L1"
28718..28742
/rpt_family="(T)n"
28735..28771
/notes="match to EST AA954400 (NID:g3118095) oo01f01.s1"
28747..28826
/notes="match to EST AW006057 (NID:g5854835) w281d04.xl"
29022..29308
/rpt_family="L1"
29384..30105
/notes="match to EST BF680971 (NID:g11954866)"
29687..30351
/notes="match to EST BG429938 (NID:g13336444)"

Query Match 94.8%; Score 1359.8; DB 9; Length 175081;
Best Local Similarity 99.3%; Pred. No. 7.8e-216;
Matches 1391; Conservative 5; Mismatches 2; Indels 3; Gaps 3;

QY 1 CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTCACCTCCATTCATGTTGA 60
DB 58149 CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTCACCTCCATTCATGTTGA 58090

QY 61 TTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTTAATTTTATATTTGAAAGTCAGTGC 120
DB 58089 TTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTTAATTTTATATTTGAAAGTCAGTGC 58030

QY 121 AAAATTATGAATAGGATATCTAATAAATACAAAGTAATAACAAAAGTCAAAGCAGTGT 180
DB 58029 AAAATTATGAATAGGATATCTAATAAATACAAAGTAATAACAAAAGTCAAAGCAGTGT 57970

QY 181 CTAATAAATAATCTCGGTTCTTAAAAATTTTAAAAATTTATC-TTGAAATAGTTTC 239
DB 57969 CTAATAAATAATCTCGGTTCTTAAAAATTTTAAAAATTTATCTTTGAAATAGTTTC 57910

QY 240 TTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCAATAA 299
DB 57909 TTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCAATAA 57850

QY 300 ACNAATTGCTATTAATGCMGA-GTGGTAATATACAGAAATTTATCAGGCATTTACCAAG 358
DB 57849 ACNAATTGCTATTAATGCMGA-GTGGTAATATACAGAAATTTATCAGGCATTTACCAAG 57790

QY 359 TCTAGGCACATATAGGAATGACGACTCAGAAATGTTCAATGTAGTGTGATGCTTG 418
DB 57789 TCTAGGCACATATAGGAATGACGACTCAGAAATGTTCAATGTAGTGTGATGCTTG 57730

QY 419 TAAGTAGGGAGCTTTATTACACATAGTAGTATGTTTCTCTAATGCTGTSTCAATGCT 478
DB 57729 TAAGTAGGGAGCTTTATTACACATAGTAGTATGTTTCTCTAATGCTGTSTCAATGCT 57670

QY 479 GGCCTTTGGCTACCTGCTACTTCCSATTATGGCAGGCCCATTCAGTCTTGAGTTTCTTCT 538
DB 57669 GGCCTTTGGCTACCTGCTACTTCCSATTATGGCAGGCCCATTCAGTCTTGAGTTTCTTCT 57610

QY 539 CTGGACACCTTATGCTCTGAATCATGAGCGGCTGATTCAATTCGTGATTTGGGTAGA 598
DB 57609 CTGGACACCTTATGCTCTGAATCATGAGCGGCTGATTCAATTCGTGATTTGGGTAGA 57550

QY 599 AAGCAGTATGTTTGTGCTGACATTAAAGATGTAGTATAGATAGGTTTAGCTTTAAGTGT 658
DB 57549 AAGCAGTATGTTTGTGCTGACATTAAAGATGTAGTATAGATAGGTTTAGCTTTAAGTGT 57490

QY 659 ATGTTTTTATACCTTTAAATAGAAATATAACCTTTTAAAGTATTCACCTCTCTCCCCA 718
DB 57489 ATGTTTTTATACCTTTAAATAGAAATATAACCTTTTAAAGTATTCACCTCTCTCCCCA 57430

QY 719 GCCTATCTCAAACTGGTGAATATATGGAGAGATCTTGAAGAAGTAATAAATCAACCTTCA 778
DB 57429 GCCTATCTCAAACTGGTGAATATATGGAGAGATCTTGAAGAAGTAATAAATCAACCTTCA 57370
```



```

* 9518 12107: gap of unknown length
* 9618 12107: contig of 2490 bp in length
* 12108 12207: gap of unknown length
* 12208 12207: contig of 2807 bp in length
* 15014 15014: gap of unknown length
* 15015 15114: gap of unknown length
* 15115 17831: contig of 2717 bp in length
* 17832 17931: gap of unknown length
* 17933 20480: contig of 2549 bp in length
* 20481 20580: gap of unknown length
* 20581 22700: contig of 2120 bp in length
* 22701 22800: gap of unknown length
* 22801 24909: contig of 2109 bp in length
* 24910 25009: gap of unknown length
* 25010 27313: contig of 2304 bp in length
* 27314 27413: gap of unknown length
* 27414 29798: contig of 2385 bp in length
* 29799 29898: gap of unknown length
* 29899 33466: contig of 3568 bp in length
* 33467 33566: gap of unknown length
* 33567 36679: contig of 3113 bp in length
* 36680 36779: gap of unknown length
* 36780 40472: contig of 3693 bp in length
* 40473 40572: gap of unknown length
* 40573 42683: contig of 2111 bp in length
* 42684 42783: gap of unknown length
* 42784 46319: contig of 3536 bp in length
* 46320 46419: gap of unknown length
* 46420 49507: contig of 3088 bp in length
* 49508 49607: gap of unknown length
* 49608 53212: contig of 3605 bp in length
* 53213 53312: gap of unknown length
* 53313 57267: contig of 3955 bp in length
* 57268 57367: gap of unknown length
* 57368 60444: contig of 3077 bp in length
* 60445 60544: gap of unknown length
* 60545 65859: contig of 5315 bp in length
* 65860 65939: gap of unknown length
* 65940 69038: contig of 3079 bp in length
* 69039 69138: gap of unknown length
* 69139 73445: contig of 4307 bp in length
* 73446 73545: gap of unknown length
* 73546 78107: contig of 4562 bp in length
* 78108 78207: gap of unknown length
* 78209 82504: contig of 4297 bp in length
* 82505 82604: gap of unknown length
* 82605 86976: contig of 4372 bp in length
* 86977 91315: contig of 4239 bp in length
* 91316 91415: gap of unknown length
* 91416 95084: contig of 3669 bp in length
* 95085 95184: gap of unknown length
* 95185 102348: contig of 7164 bp in length
* 102349 102448: gap of unknown length
* 102449 105973: contig of 3525 bp in length
* 105974 106073: gap of unknown length
* 106074 112891: contig of 6818 bp in length
* 112892 112892: gap of unknown length
* 112893 116221: contig of 3230 bp in length
* 116222 116321: gap of unknown length
* 116322 119798: contig of 3477 bp in length
* 119799 119799: gap of unknown length
* 119800 124870: contig of 4972 bp in length
* 124871 124970: gap of unknown length
* 124971 130790: contig of 5820 bp in length
* 130791 130890: gap of unknown length
* 130891 134639: contig of 3749 bp in length
* 134640 134739: gap of unknown length
* 134740 140637: contig of 5898 bp in length
* 140638 140738: gap of unknown length
* 140739 145541: contig of 4804 bp in length
* 145542 145641: gap of unknown length
* 145642 155157: contig of 9416 bp in length
* 155158 155157: gap of unknown length

* 155158 164357: contig of 9200 bp in length
* 164358 164457: gap of unknown length
* 164458 174892: contig of 10435 bp in length
* 174893 174992: gap of unknown length
* 174993 184400: contig of 9408 bp in length
* 184401 184500: gap of unknown length
* 184501 198385: contig of 13885 bp in length
* 198386 198485: gap of unknown length
* 198486 216436: contig of 17951 bp in length
* 216437 216536: gap of unknown length
* 216537 232916: contig of 16380 bp in length
* 232917 233016: gap of unknown length
* 233017 255437: contig of 22421 bp in length
* 255438 255537: gap of unknown length
* 255538 281662: contig of 26125 bp in length.

FEATURES             Location/Qualifiers
     source            1..281662
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="12"
                        /clone="RP11-19D19"

ORIGIN
Query Match      41.2%; Score 591; DB 2; Length 281662;
Best Local Similarity 98.1%; Pred. No. 9.8e-89;
Matches 617; Conservative 2; Mismatches 7; Indels 3; Gaps 2;

QY 773 CCTTCACTGCTCCACTCCAGGTGAATCCGCGCCACTCCACTGACCTAGTAGAATTGTAA 832
DB 33860 CCTCACCGCTCCCTCCACGTGAATCCGCGCCACTCCACTGACCTAGTAGAATTGTAA 33919

QY 833 TTATAATAC--TTACCTCTTATTTCTGAAATCAGTTGTGAACCTGTGCTTTATGTTTCAGAR 890
DB 33920 TTTAATACGTTAGCGTTCTATTTCTGAATCAGTTGTGAACCTGTGCTTTATGTTTCAGAR 33979

QY 891 GTTTAAGAACCTCMGTGAAATTCATTTTTTAAATCTGCTATTCTGAGAAGCAATTTGAATGA 950
DB 33980 GTTTAAGAACCTCAGTGAATTCATTTTTTAAATCTGCTATTCTGAGAAGCAATTTGAATGA 34039

QY 951 ATTCTTAAACAGAAGACTCATCTGTAGCTGTTTGTGCTGATCTCTATGAGCCCAATAGGGT 1010
DB 34040 ATTCTTAAACAGAAGACTCATCTGTAGCTGTTTGTGCTGATCTCTATGAG--CCCATAGGGT 34098

QY 1011 TCTGTGCTTAGCATTAACAAATAAGGTTTATAGGTAAGCCCAATGCTATTATTTT 1070
DB 34099 TCTGTGCTTAGCATTAACAAATAAGGTTTATAGGTAAGCCCAATGCTATTATTTT 34158

QY 1071 TGCATGGAGGCGTTTAAAAATTTGTCCTCTTTTTCATATTTTATTCATATTTCAATTTATGG 1130
DB 34159 TGCATGGAGGCGTTTAAAAATTTGTCCTCTTTTTCATATTTTATTCATATTTATGG 34218

QY 1131 TTTGTAACTGCTTTTATAGGGAGTAAATATATGCTTTTAAATTTAGTTTGGGGGGAATAAT 1190
DB 34219 TTTGTAACTGCTTTTATAGGGAGTAAATATATGCTTTTAAATTTAGTTTGGGGGGAATAAT 34278

QY 1191 TGTGCAAGAGGATAATTTAAATTTACGTGCTCTCTGTTTATTCAGATAAAGAGAGAGAGACT 1250
DB 34279 TGTGCAAGAGGATAATTTAAATTTACGTGCTCTCTGTTTATTCAGATAAAGAGAGAGACT 34338

QY 1251 AGCTGCATATTCAGAGGTTGTACCTTAAACATTTGGTGAACATTTTCTTAAGATTTTCA 1310
DB 34339 AGCTGCATATTCAGAGGTTGTACCTTAAACATTTGGTGAACATTTTCTTAAGATTTTCA 34398

QY 1311 AAAGGAATATGCTAAATTTGAGNAATCATACCACTGTCCTTAACCTTGGTAAACAACTGT 1370
DB 34399 AAAGGAATATGCTAAATTTGAGNAATCATACCACTGTCCTTAACCTTGGTAAACAACTGT 34458

QY 1371 TCTTAATAAAGTATTATTAATGATTTTAA 1399
DB 34459 TCTTAATAAAGTATTATTAATGATTTTAA 34487

```

```

RESULT 7
AC117374/c
LOCUS
DEFINITION Homo sapiens chromosome 12 clone RP11-19D19, *** SEQUENCING IN
PROGRESS ***, 48 unordered pieces.
ACCESSION AC117374 AC011694
VERSION AC117374.1 GI:20127352
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 281662)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Aisbrooks,S.B., Amaratunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroli,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabris,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hoggues,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsone,B., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louisseg,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,N.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickett,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojebokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sison,I.,
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Ugmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 281662)
Worley,K.C.
Direct Submission
Submitted (10-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 281662)
Worley,K.C.
Direct Submission
Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:10045385.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Web contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: HCND
Center clone name: RP11-19D19
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 273063 bases at least Q40
Consensus quality: 294336 bases at least Q30
Consensus quality: 306942 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 48 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1
2464: contig of 2464 bp in length
2564: gap of unknown length
2565: contig of 2039 bp in length
4603: gap of unknown length
4703: gap of unknown length
7113: contig of 2410 bp in length
7213: gap of unknown length
9517: contig of 2304 bp in length
9617: gap of unknown length
12107: contig of 2490 bp in length
12207: gap of unknown length
15014: contig of 2807 bp in length
15114: gap of unknown length
17831: contig of 2717 bp in length
17931: gap of unknown length
20480: contig of 2549 bp in length
20580: gap of unknown length
22700: contig of 2120 bp in length
22800: gap of unknown length
22801: contig of 2109 bp in length
25009: gap of unknown length
27313: contig of 2304 bp in length
27413: gap of unknown length
29798: contig of 2385 bp in length
29898: gap of unknown length
33466: contig of 3568 bp in length
33566: gap of unknown length
36679: contig of 3113 bp in length
36779: gap of unknown length
40472: contig of 3693 bp in length
40572: gap of unknown length
42683: contig of 2111 bp in length
42783: gap of unknown length
46319: contig of 3536 bp in length
46419: gap of unknown length
49507: contig of 3088 bp in length
49607: gap of unknown length
53212: contig of 3605 bp in length
53312: gap of unknown length
57267: contig of 3955 bp in length
57367: gap of unknown length
60444: contig of 3077 bp in length
60544: gap of unknown length
65859: contig of 5315 bp in length
65959: gap of unknown length
69038: contig of 3079 bp in length
69138: gap of unknown length
73445: contig of 4307 bp in length
73545: gap of unknown length
78107: contig of 4562 bp in length
78207: gap of unknown length
82504: contig of 4297 bp in length
82604: gap of unknown length

```



```
* 82605 86976: contig of 4372 bp in length
* 86977 87076: gap of unknown length
* 87077 91315: contig of 4239 bp in length
* 91315 91415: gap of unknown length
* 91415 95084: contig of 3669 bp in length
* 95084 95184: gap of unknown length
* 95184 102348: contig of 7164 bp in length
* 102348 102448: gap of unknown length
* 102448 105973: contig of 3525 bp in length
* 105973 106073: gap of unknown length
* 106073 112891: contig of 6818 bp in length
* 112891 112892: gap of unknown length
* 112892 112992: contig of 3230 bp in length
* 112992 116321: gap of unknown length
* 116321 119798: contig of 3477 bp in length
* 119798 119898: gap of unknown length
* 119898 124870: contig of 4972 bp in length
* 124870 124970: gap of unknown length
* 124970 130790: contig of 5820 bp in length
* 130790 130890: gap of unknown length
* 130890 134639: contig of 3749 bp in length
* 134639 134640: gap of unknown length
* 134640 140637: contig of 5898 bp in length
* 140637 140638: gap of unknown length
* 140638 140738: contig of 4804 bp in length
* 140738 145541: gap of unknown length
* 145541 145641: gap of unknown length
* 145641 155057: gap of 9416 bp in length
* 155057 155157: gap of unknown length
* 155157 164357: contig of 9200 bp in length
* 164357 164457: gap of unknown length
* 164457 174832: contig of 10435 bp in length
* 174832 174992: gap of unknown length
* 174992 184400: contig of 9408 bp in length
* 184400 184500: gap of unknown length
* 184500 198385: contig of 13885 bp in length
* 198385 198485: gap of unknown length
* 198485 216437: contig of 17951 bp in length
* 216437 216537: gap of unknown length
* 216537 232916: contig of 16380 bp in length
* 232916 233016: gap of unknown length
* 233016 255437: contig of 22421 bp in length
* 255437 255438: gap of unknown length
* 255438 281662: contig of 26125 bp in length.
* 281662 281662: contig of 26125 bp in length.

FEATURES
    source
    1..281662
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="12"
        /clone="RP11-19D19"

ORIGIN
    Query Match      38.7%; Score 555; DB 2; Length 281662;
    Best Local Similarity 98.6%; Pred. No. 8.8e-83;
    Matches 577; Conservative 3; Mismatches 3; Indels 2; Gaps 2;

QY 1 CATTAACTCTTTTTCGGGAATAGTATGATATTTCAATGFCACCTCCATCATCTTGA 60
DB 146226 CATTAACTCTTTTATCGGAATAGTATGATATTTTCAGTGTCACTCCATTCATCTTGA 146167
QY 61 TTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTTTATATTGAAAGTCAGTGC 120
DB 146166 TTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTTTATATTGAAAGTCAGTGC 146107
QY 121 AAAAAATGATAGGATATATCTAAATAAATAACAAAGTAATAACAAAAGTCAAGCAGTGT 180
DB 146106 AAAAAATGATAGGATATATCTAAATAAATAACAAAGTAATAACAAAAGTCAAGCAGTGT 146047
QY 181 CTAATAAAAAATCTCGGTTCCTTAAAAATATTTTAAAAATTTATC-TTGAAATAGTTTTTC 239
DB 146046 CTAATAAAAAATCTCGGTTCCTTAAAAATATTTTAAAAATTTATCTTTGAAATAGTTTTTC 145987
QY 240 TTAGATTATCTCAGGATATGACAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 299
```

```
Db 145986 TTAGATTATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 145927
QY 300 ACAAATTGCTCTATTAAATGCAMGA-CTGGTAATATACAGAAATTTATCAGGCATTACCAAG 358
DB 145926 ACAAATTGCTCTATTAAATGCAMGACGTGGTAATATACAGAAATTTATCAGGCATTACCAAG 145867
QY 359 TCTAGSCACATATAGGAATGCAGCACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTG 418
DB 145866 TCTAGSCACATATAGGAATGCAGCACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTG 145807
QY 419 TAAGGTAGGGAGCTTATTTCAGACATAGTAGTAGTTTCTTAATGCTGTSTCAATGCT 478
DB 145806 TAAGGTAGGGAGCTTATTTCAGACATAGTAGTAGTTTCTTAATGCTGTSTCAATGCT 145747
QY 479 GGCCTTTGGCTACCTGTACTTCCSATTATGGCAGCCCAATCAGTCTTCAGTTTTCTTCT 538
DB 145746 GGCCTTTGGCTACCTGTACTTCCSATTATGGCAGCCCAATCAGTCTTCAGTTTTCTTCT 145687
QY 539 CTGGACACCTTATGCTCTGAAATCATGAGCGAGGCTGATTCAATT 583
DB 145686 CTGGACACCTATGCTCTGAAATCATGAGCGAGGCTGATTCAATT 145642

RESULT 8
AX408738/c
LOCUS AX408738 448 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1385 from Patent WO0229103.
ACCESSION AX408738
VERSION AX408738.1 GI:21441443
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
    1
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 1385 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
    Location/Qualifiers
    source
    1..448
        /organism="Homo sapiens"
        /mol_type="unassigned DNA"
        /db_xref="taxon:9606"
        /note="EMBL/GenBank Accession No. AA491000"

ORIGIN
    Query Match      29.7%; Score 425.8; DB 6; Length 448;
    Best Local Similarity 98.2%; Pred. No. 7.5e-61;
    Matches 441; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
```

Db 149 AGAGTTGACCTTAACATTTGGTGAACATTTTCTAAGATTTTCARAAAGATATGTGT 90
Qy 1325 AAATTGAGAAATCATACCACTGCTCTAATCTGGTAAACAACACTGTTCTTAAATAAGTA 1384
Db 89 AAATTGAGAAATCATACCACTGCTCTAATCTGGTAAACAACACTGTTCTTAAATAAGTA 30
Qy 1385 TTTTAATGATTTTAAAAAAGAAAAA 1413
Db 29 TTTGATGATGTTTAAAAAAGAGAGAA 1
RESULT 9
AX898296
LOCUS AX898296 395 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 14159 from Patent EP1033401.
ACCESSION AX898296
VERSION AX898296.1 GI:40053209
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Dumas Milne Edwards, J.B., Duclert, A. and Giordano, J.Y.
TITLE Expressed sequence tags and encoded human proteins
JOURNAL Patent: EP 1033401-A 14159 06-SEP-2000;
Genset (FR)
FEATURES
source Location/Qualifiers
1 .395
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 24.9%; Score 356.8; DB 6; Length 395;
Best Local Similarity 95.7%; Pred. No. 2e-49;
Matches 378; Conservative 8; Mismatches 6; Indels 3; Gaps 2;
Qy 58 TGATTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAAATTTATATTGAAAGTCAG 117
Db 1 TGATTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAAATTTATATTGAAAGTCAG 60
Qy 118 TGCATAAATTATGATAGGATATCTAATAATAACAAGTAATAACAAGTCAAAAGCAGT 177
Db 61 TGCATAAATTATGATAGGATATCTAATAATAACAAGTAATAACAAGTCAAAAGCAGT 120
Qy 178 GTTCTAAATAAATAATTCGGGTCCCTTAAATAATTTTAAATTTT--ATCTTGAATAGT 235
Db 121 GTTCTAAATAAATAATTCGGGTCCCTTAAATAATTTTAAATTTTATCTTTGAAATAGT 180
Qy 236 TTTCTTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCA 295
Db 181 NWTCTTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCA 240
Qy 296 TTAACAATAATGCTATTAATAATGCMGA-GTGTAATATACAGAAATTTATCAGGCATTAC 354
Db 181 NWTCTTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCA 240
Qy 296 TTAACAATAATGCTATTAATAATGCMGA-GTGTAATATACAGAAATTTATCAGGCATTAC 354
Db 241 TTAACAATAATGCTATTAATAATGCMGA-GTGTAATATACAGAAATTTATCAGGCATTAC 300
Qy 355 CAAGTCTAGGCACATATAGGAATGCGAGCTCAGAAATGCTTCAATGTAGTGTGATG 414
Db 301 CAAGTCTAGGCACATATAGGAATGCGAGCTCAGAAATGCTTCAATGTAGTGTGATG 414
Qy 415 CTTGTAAGGTAGGGAGCTTTATTCAGACATAGTAG 449
Db 361 CTTGTAAGGTAGGGAGCTTTATTCAGACATAGTAG 395
RESULT 10
BD033829
LOCUS BD033829 395 bp DNA linear PAT 27-AUG-2002
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD033829

VERSION BD033829.1 GI:22575571
KEYWORDS JP 2001269182-A/10075.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 395)
AUTHORS Edwards, J.B.D.M., Duclair, E. and Jordan, J.Y.
TITLE Sequence tag and encoded human protein
JOURNAL Patent: JP 2001269182-A 10075 02-OCT-2001;
GENSET
COMMENT OS Homo sapiens (human)
PN JP 2001269182-A/10075
PD 02-OCT-2001
PF 24-FEB-2000 JP 2000118773
PR 26-FEB-1999 US 60/122487
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES
PI JORDAN
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10
PC C12P21/02, C12P21/08, C12Q1/68//G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC
FH Key Location/Qualifiers.
FEATURES
source Location/Qualifiers
1 .395
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 24.9%; Score 356.8; DB 6; Length 395;
Best Local Similarity 95.7%; Pred. No. 2e-49;
Matches 378; Conservative 8; Mismatches 6; Indels 3; Gaps 2;
Qy 58 TGATTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAAATTTATATTGAAAGTCAG 117
Db 1 TGATTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAAATTTATATTGAAAGTCAG 60
Qy 118 TGCATAAATTATGATAGGATATCTAATAATAACAAGTAATAACAAGTCAAAAGCAGT 177
Db 61 TGCATAAATTATGATAGGATATCTAATAATAACAAGTAATAACAAGTCAAAAGCAGT 120
Qy 178 GTTCTAAATAAATAATTCGGGTCCCTTAAATAATTTTAAATTTT--ATCTTGAATAGT 235
Db 121 GTTCTAAATAAATAATTCGGGTCCCTTAAATAATTTTAAATTTTATCTTTGAAATAGT 180
Qy 236 TTTCTTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCA 295
Db 181 NWTCTTAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCA 240
Qy 296 TTAACAATAATGCTATTAATAATGCMGA-GTGTAATATACAGAAATTTATCAGGCATTAC 354
Db 241 TTAACAATAATGCTATTAATAATGCMGA-GTGTAATATACAGAAATTTATCAGGCATTAC 300
Qy 355 CAAGTCTAGGCACATATAGGAATGCGAGCTCAGAAATGCTTCAATGTAGTGTGATG 414
Db 301 CAAGTCTAGGCACATATAGGAATGCGAGCTCAGAAATGCTTCAATGTAGTGTGATG 414
Qy 415 CTTGTAAGGTAGGGAGCTTTATTCAGACATAGTAG 449
Db 361 CTTGTAAGGTAGGGAGCTTTATTCAGACATAGTAG 395
RESULT 11
AC121881
LOCUS AC121881 184359 bp DNA linear ROD 11-NOV-2003
DEFINITION Mus musculus BAC clone RP24-13418 from chromosome 8, complete
sequence.
ACCESSION AC121881
VERSION AC121881.3 GI:23462961
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)

```

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
Harris,A., Meyer,R. and Spalding,L.
TITLE
The sequence of Mus musculus BAC clone RP24-134L8
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 184359)
TITLE
Sequencing of Mus musculus
JOURNAL
REFERENCE
AUTHORS
3 (bases 1 to 184359)
TITLE
Unpublished (2001)
JOURNAL
REFERENCE
AUTHORS
McPherson,J.D. and Waterston,R.H.
TITLE
Direct Submission
JOURNAL
REFERENCE
AUTHORS
Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
TITLE
Parkway, St. Louis, MO 63108, USA
JOURNAL
REFERENCE
AUTHORS
4 (bases 1 to 184359)
TITLE
McPherson,J.D. and Waterston,R.H.
TITLE
Direct Submission
JOURNAL
REFERENCE
AUTHORS
Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
TITLE
Parkway, St. Louis, MO 63108, USA
JOURNAL
REFERENCE
AUTHORS
5 (bases 1 to 184359)
TITLE
McPherson,J.D. and Waterston,R.H.
TITLE
Direct Submission
JOURNAL
REFERENCE
AUTHORS
Submitted (02-OCT-2002) Genome Sequencing Center, 4444 Forest Park
TITLE
Parkway, St. Louis, MO 63108, USA
JOURNAL
REFERENCE
AUTHORS
6 (bases 1 to 184359)
TITLE
Wilson,R.
TITLE
Direct Submission
JOURNAL
REFERENCE
AUTHORS
Submitted (11-NOV-2003) Department of Genetics, Washington
TITLE
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL
COMMENT
On Oct 2, 2002 this sequence version replaced gi:22475821.
----- Genome Center
Center: Washington University Genome Sequencing Center
Web site: http://genome.wustl.edu
Contact: submissions@genome.wustl.edu
----- Summary Statistics
Center project name: M_BB0134L08
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is
overlapped by AC117196.
Location/Qualifiers
1. .184359
/organism="Mus musculus"
/mol_type="genomic DNA"

```

```

/db_xref="taxon:10090"
/chromosome="8"
/map="8"
/clone="RP24-134L8"
/clone_lib="RPCI-24"
289. .527
/rpt_family="B4"
1994. .2232
/rpt_family="ERVK"
2383. .3887
/rpt_family="L1"
3888. .4101
/rpt_family="ERVK"
4327. .4514
/rpt_family="MaLR"
4652. .4817
/rpt_family="RMER15"
5330. .5642
/rpt_family="MaLR"
6281. .6619
/rpt_family="RMER10B"
6759. .7020
/rpt_family="B4"
7183. .7363
/rpt_family="B2"
7436. .7681
/rpt_family="MaLR"
7701. .7824
/rpt_family="MaLR"
9897. .10028
/rpt_family="Alu"
10171. .10305
/rpt_family="Alu"
10314. .10449
/rpt_family="B4"
11501. .11617
/rpt_family="Alu"
12345. .12438
/rpt_family="BCI_MM"
12471. .12629
/rpt_family="B4"
12684. .12778
/rpt_family="B4"
12914. .13275
/rpt_family="L1"
13283. .14578
/rpt_family="L1"
14627. .15630
/rpt_family="L1"
15739. .15857
/rpt_family="ERV1"
16247. .16375
/rpt_family="Alu"
16501. .16615
/rpt_family="Alu"
16655. .16713
/rpt_family="Alu"
17787. .17878
/rpt_family="Alu"
17950. .18029
/rpt_family="Alu"
18461. .18609
/rpt_family="B4"
18816. .18908
/rpt_family="Alu"
19317. .19666
/rpt_family="MaLR"
19739. .19928
/rpt_family="B2"
20490. .20679
/rpt_family="B2"
21147. .21374
/rpt_family="L1"

```


clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J., and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTD-2195F2, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-425P5; actual end is at base position 212633 of RP11-425P5.

FEATURES

source

1. -212827
/organism="Homo sapiens"
/mol_type="Genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"

/clone="RP11-425P5"
/clone_lib="RPCI-11"

repeat_region

1623..1696

/rpt_family="MIR"

repeat_region

1698..2006

/rpt_family="Alu"

repeat_region

2024..2337

/rpt_family="Alu"

repeat_region

3006..3093

/rpt_family="L1"

repeat_region

3453..3552

/rpt_family="L2"

repeat_region

4924..5046

/rpt_family="ERV1"

misc_feature

5318..5693

/note="similar to EST BF996284 (NID:g12402607)"

repeat_region

5580..5845

/rpt_family="Alu"

repeat_region

5925..6277

/rpt_family="MaLR"

misc_feature

6842..7041

/note="similar to EST BF369507 (NID:g11331532)"

repeat_region

7384..7661

/rpt_family="Alu"

repeat_region

8029..8048

/rpt_family="Alu"

repeat_region

8379..8678

/rpt_family="Alu"

misc_feature

8560..8762

/note="similar to EST AA699398 (NID:g2702592) z140a03.s1"

repeat_region

8720..9031

/rpt_family="Alu"

misc_feature

9083..9560

/note="similar to EST AA633603 (NID:g2556817) ae66b02.s1"

repeat_region

9598..9899

/rpt_family="Alu"

misc_feature

9871..10098

/note="similar to EST AA699398 (NID:g2702592) z140a03.s1"

repeat_region

10746..11044

/rpt_family="L1"

repeat_region

11075..11101

/rpt_family="AT-rich"

repeat_region

11102..11372

/rpt_family="Alu"

repeat_region

11411..11465

/rpt_family="GA-rich"

repeat_region

11552..11644

/rpt_family="L2"

repeat_region

11685..12193

/rpt_family="L1"

repeat_region

12194..12486

/rpt_family="Alu"

repeat_region

12487..13003

/rpt_family="L1"

repeat_region

13004..13305

/rpt_family="Alu"

repeat_region

13306..13501

/rpt_family="L1"

repeat_region

13503..13892

/rpt_family="L1"

misc_feature

13893..14224

/note="similar to EST BE061102 (NID:g8405752)"

repeat_region

14244..14544

/rpt_family="Alu"

repeat_region

14546..14828

/rpt_family="Alu"

repeat_region

15091..15189

/rpt_family="MIR"

repeat_region

15672..16449

/rpt_family="ACHobo"

repeat_region

16640..16844

/rpt_family="Alu"

repeat_region

16917..16959

/rpt_family="ACHobo"

repeat_region

16960..17287

/rpt_family="Alu"

repeat_region

17288..17572

/rpt_family="ACHobo"

misc_feature

17610..17938

/note="similar to EST BF996427 (NID:g12402750)"

repeat_region

17793..18068

/rpt_family="Alu"

repeat_region

18251..18537

/rpt_family="Alu"

misc_feature

18485..18884

/note="similar to EST BE062347 (NID:g8406997)"

misc_feature

18661..18809

/note="similar to EST BG186544 (NID:g13708231)"

repeat_region

18957..19160

/rpt_family="L1"

repeat_region

19314..19528

/rpt_family="L1"

repeat_region

19619..20135

/rpt_family="Alu"

repeat_region

21101..21269

/rpt_family="L1"

repeat_region

21882..22203

/rpt_family="L1"

misc_feature

22429..22797

/note="similar to EST BF993478 (NID:g12399801)"

repeat_region

22717..23007

/rpt_family="L1"

misc_feature

22993..23191


```
* 20722 21529: contig of 808 bp in length
* 21530 21629: gap of 100 bp
* 21630 22393: contig of 764 bp in length
* 22394 22493: gap of 100 bp
* 22494 23260: contig of 767 bp in length
* 23261 23360: gap of 100 bp
* 23361 24148: contig of 788 bp in length
* 24149 24248: gap of 100 bp
* 24249 25036: contig of 788 bp in length
* 25037 25136: gap of 100 bp
* 25137 25906: contig of 770 bp in length
* 25907 26006: gap of 100 bp
* 26007 26773: contig of 767 bp in length
* 26774 26873: gap of 100 bp
* 26874 27630: contig of 757 bp in length
* 27631 27730: gap of 100 bp
* 27731 28554: contig of 824 bp in length
* 28555 29483: contig of 829 bp in length
* 29484 29583: gap of 100 bp
* 29584 30372: contig of 789 bp in length
* 30373 30472: gap of 100 bp
* 30473 31230: contig of 758 bp in length
* 31231 31330: gap of 100 bp
* 31331 32132: contig of 802 bp in length
* 32133 32232: gap of 100 bp
* 32233 33005: contig of 773 bp in length
* 33006 33105: gap of 100 bp
* 33106 33862: contig of 757 bp in length
* 33863 34729: gap of 100 bp
* 34730 34829: gap of 100 bp
* 34830 35597: contig of 768 bp in length
* 35598 35697: gap of 100 bp
* 35698 36479: contig of 782 bp in length
* 36480 36579: gap of 100 bp
* 36580 37350: contig of 771 bp in length
* 37351 37450: gap of 100 bp
* 37451 38223: contig of 773 bp in length
* 38224 38323: gap of 100 bp
* 38324 38977: contig of 654 bp in length
* 38978 39077: gap of 100 bp
* 39078 39863: contig of 786 bp in length
* 39864 39963: gap of 100 bp
* 39964 40757: contig of 794 bp in length
* 40758 40857: gap of 100 bp
* 40858 41620: contig of 763 bp in length
* 41621 41720: gap of 100 bp
* 41721 42451: contig of 731 bp in length
* 42452 42551: gap of 100 bp
* 42552 43301: contig of 750 bp in length
* 43302 43401: gap of 100 bp
* 43402 44183: contig of 782 bp in length
* 44184 44283: gap of 100 bp
* 44284 45037: contig of 754 bp in length
* 45038 45137: gap of 100 bp
* 45138 45897: contig of 760 bp in length
* 45898 45997: gap of 100 bp
* 45998 46781: contig of 784 bp in length
* 46782 46881: gap of 100 bp
* 46882 47677: contig of 796 bp in length
* 47678 47777: gap of 100 bp
* 47778 48535: contig of 758 bp in length
* 48536 48635: gap of 100 bp
* 48636 49385: contig of 750 bp in length
* 49386 49485: gap of 100 bp
* 49486 50288: contig of 803 bp in length
* 50289 50388: gap of 100 bp
* 50389 51159: contig of 771 bp in length
* 51160 51259: gap of 100 bp
* 51260 51995: contig of 736 bp in length
* 51996 52095: gap of 100 bp
* 52096 52895: contig of 770 bp in length
```

```
* 52866 52965: gap of 100 bp
* 52966 53738: contig of 773 bp in length
* 53739 53838: gap of 100 bp
* 53839 54621: contig of 783 bp in length
* 54622 54721: gap of 100 bp
* 54722 55479: contig of 758 bp in length
* 55480 55579: gap of 100 bp
* 55580 56361: contig of 782 bp in length
* 56362 56461: gap of 100 bp
* 56462 57212: contig of 751 bp in length
* 57213 57312: gap of 100 bp
* 57313 58070: contig of 758 bp in length
* 58071 58170: gap of 100 bp
* 58171 58945: contig of 775 bp in length
* 58946 59045: gap of 100 bp
* 59046 59771: contig of 726 bp in length

Query Match 5.3%; Score 76.6; DB 2; Length 72832;
Best Local Similarity 36.3%; Pred. No. 0.0015;
Matches 154; Conservative 0; Mismatches 268; Indels 2; Gaps 1;

Qy 1010 TTCTGTGCTTAGCATTAACAAATAAGGTTTATAGGTAAGCCCAATGTTATTTT 1069
Db 42355 TTTTNNNNNTTTTATTATATAAAAAATAAATANTTAATTATATTNNTTAA 42296
Qy 1070 TTGCATGGAGGCTTTAAAAATTTTGCTCTTTTTCATTTTTCATATTTCAT 1129
Db 42295 NNNNNNNNTTTTNTTANTTTTNTTNNNNNTTNNNNNTTNNNNNTTNNNN 42238
Qy 1130 GTTTGTAACGCTTTTATAGGAGATAATATATAGTTATATAATAGTTTGG 1189
Db 42237 TATNNNTANNNTTTTNNNTTAAANATAATTTTNTTNNNTTNNNNNTTN 42178
Qy 1190 TTGTGCAAGAGGATAATTTAAATTACGTCCTCTGTTTTCAGATAAAGAG 1249
Db 42177 NNNNNNANTNTNANNNNNNNNAATTTTNNNNNTTTTNNNNNTTAAAAA 42118
Qy 1250 TACGCTGCATATCAAGAGTTGTACCTTAACATTGGTGAACATTTTCTA 1309
Db 42117 NANNNTNATTTTAAATNAAATTTNNNNNAAATTTNNNNNNTNNNNNNA 42058
Qy 1310 AAAAGGAATATGTAAATTTAGAAATCATACCACTGCTCTAACTTGGTAA 1369
Db 42057 AAAANNTTANTTTTTTTTNTANNNANNAANNAANCCNNANATAAAA 41998
Qy 1370 TTCTTAAATTAAGTATTATGATTTTAAAAAATAAAAAAATAAAAAA 1429
Db 41997 NCCNCAAAAAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNA 41938
Qy 1430 AAAA 1433
Db 41937 TATA 41934
```

RESULT 15
AC116984 1/c
WPCOMMENT

Sequence split into 6 fragments LOCUS AC116984 Accession AC116984

Fragment Name	Begin	End
AC116984_0	1	110000
AC116984_1	100001	210000
AC116984_2	200001	310000
AC116984_3	300001	410000
AC116984_4	400001	510000
AC116984_5	500001	541399

Continuation (2 of 6) of AC116984 from base 100001 (AC116984 Dictyostelium discoideum)

Query Match 5.1%; Score 73.2; DB 3; Length 110000;
Best Local Similarity 48.8%; Pred. No. 0.0051;
Matches 198; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

Qy 1029 AAAATAAGTTTATAGGTAAGCCCAATGTTATTAATTTTTTTTGGCATGGAGGCTTAAA 1088

```
Db 10330 AAAATGATCACCTTCTTTTCCTTTTAAATCGTTTCATTTTTTTTTTATTTTTTTT 10271
QY 1089 ATTGTGCTCTTTTTCATATTTTATTCATATTCATATTCATATTCATATTCATATTCAT 1148
Db 10270 TTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTT 10211
QY 1149 GGAGATAATATATGTTATTAATATGTTATGTTATGTTATGTTATGTTATGTTATGTTAT 1208
Db 10210 CCTGCAAAAAATTTGTGGGCAACCTCTTAGTATTTAAATTTTAAAGTTTCGGTTTT 10151
QY 1209 TAATTTACGTCTTCTGTTATTCAGATTAAGAGAGAGACTACGCTGCATATTCAGAG 1268
Db 10150 TAGTGGCCATTCCTCTCACACATGCGAAAAAATAAATAAATAAATAAATAAATAAATA 10091
QY 1269 TTGTACCTTAAACATTCGTGAAACATTTTCTAAGATTTTCAAAAGGAATATGTAAAT 1328
Db 10090 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 10031
QY 1329 TGAGAAATCATACCACTGCTCTTAATTCGTTAACTGTTAACTGTTAACTGTTAACTGTT 1388
Db 10030 ATACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 9971
QY 1389 ATGATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1434
Db 9970 ATATTTATAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 9925
```

Search completed: July 20, 2004, 18:51:00
Job time : 5507 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 15:59:46 ; Search time 583 Seconds

(without alignments)

10449.256 Million cell updates/sec

Title: US-10-799-747-12

Perfect score: 1434

Sequence: 1 cattaactcttttcttcgg.....aaaaaaaaaaaaaaaaaaaaa 1434

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1432	99.9	1434	7 ADA39771	Ada39771 Human sec
2	1432	99.9	1434	7 ACC50424	Acc50424 Human sec
3	1430.4	99.7	1434	2 AAX37452	Aax37452 Human sec
C 4	425.8	29.7	448	6 ABN94887	Abn94887 Gene #138
5	360	25.1	415	3 AAH30357	Aah30357 Human col
6	356.8	24.9	395	3 AAC10084	Aac10084 Human sec
C 7	300.4	20.9	336	2 AAT25136	Aat25136 Human gen
C 8	69.8	4.9	545	4 AAH70126	Aah70126 Human cer
C 9	69	4.8	8056	7 ABZ10246	Abz10246 Haematopo
C 10	67.8	4.7	419	7 ABX46069	Abx46069 Bovine ES
C 11	65.6	4.6	597	4 AAH71472	Aah71472 Human cer
12	63.8	4.4	8056	7 ABZ10246	Abz10246 Haematopo
13	62.6	4.4	8056	7 ABZ10100	Abz10100 Haematopo
C 14	62.6	4.4	8056	7 ABZ10100	Abz10100 Haematopo
15	62	4.3	2103	5 AAD08708	Aad08708 Lycopersi
C 16	61.2	4.3	393	7 ABX39417	Abx39417 Bovine ES
17	61.2	4.3	887	3 AAC59257	Aac59257 Human sec
18	60.8	4.2	2195	3 AAAL4072	Aaal4072 Human SPR
C 19	60.6	4.2	555	3 AAC66747	Aac66747 DNA marke
20	60.4	4.2	329	6 ABN94208	Abn94208 Gene #706
21	60.2	4.2	6076	4 AAS46664	Aas46664 Tumour su
22	60	4.2	6195	6 ABL32590	AbL32590 Human imm
C 23	60	4.2	17389	6 ABL33415	AbL33415 Human imm

C 24	59.8	4.2	6065	6 ABK31357	Abk31357 Signal tr
C 25	59.8	4.2	6065	6 ABL70580	AbL70580 Chemical
C 26	59.8	4.2	6065	6 AAS61261	Aas61261 Human gen
C 27	59.2	4.1	337	7 ABX41811	Abx41811 Bovine ES
28	59.2	4.1	17294	6 ABL32986	AbL32986 Human imm
29	59	4.1	476	8 ACH28349	Ach28349 Human adu
30	59	4.1	7352	6 ABL32370	AbL32370 Human imm
31	58.8	4.1	2976	7 ABZ10131	Abz10131 Haematopo
32	58.8	4.1	3366	4 AAF24684	Aaf24684 Nucleotid
33	58.8	4.1	3366	4 AAE24706	Aae24706 Nucleotid
34	58.8	4.1	5976	9 ADB54233	Adb54233 Pretreat
35	58.8	4.1	5976	9 ADE84171	Ade84171 Human lym
36	58.8	4.1	10365	9 ADD94068	Add94068 Human ATP
37	58.8	4.1	10381	9 ADD94069	Add94069 Human ATP
38	58.8	4.1	10423	9 ADD94070	Add94070 Human ATP
39	58.8	4.1	10442	4 AAF24680	Aaf24680 Nucleotid
40	58.8	4.1	10442	4 AAF24702	Aaf24702 Nucleotid
41	58.8	4.1	10474	4 AAF24685	Aaf24685 Nucleotid
42	58.8	4.1	10474	4 AAF24686	Aaf24686 Nucleotid
43	58.8	4.1	10474	4 AAF24708	Aaf24708 Nucleotid
44	58.8	4.1	10474	4 AAF24707	Aaf24707 Nucleotid
45	58.6	4.1	424	7 ABX46053	Abx46053 Bovine ES

ALIGNMENTS

RESULT 1

ADA39771

ID ADA39771 standard; cDNA; 1434 BP.

XX ADA39771;

XX ADA39771;

DT 20-NOV-2003 (first entry)

DE Human secreted protein encoding cDNA.

XX Human;

KW Human; secreted protein; cancer; hyperproliferative disorder;

KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;

KW anaemia; allergic reaction; asthma; cardiovascular disorder;

KW wound healing; cytostatic; immunosuppressive; neutropenic; neuroprotective;

KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;

KW vulnery; cardiac; gene therapy; ss.

XX Homo sapiens.

OS Homo sapiens.

PN WO2002102993-A2.

XX 27-DEC-2002.

PD 19-MAR-2002; 2002WO-US008123.

XX 19-MAR-2002; 2002WO-US008123.

XX 21-MAR-2001; 2001US-0277340P.

PR 19-JUL-2001; 2001US-0306171P.

PR 13-NOV-2001; 2001US-0331287P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-175238/17.

XX New human secreted proteins and nucleic acid molecules, useful for

PT preparing a diagnostic or pharmaceutical composition for diagnosing,

PT preventing or treating cancer or other hyperproliferative disorder,

PT asthma, allergies or AIDS.

XX Claim 9; SEQ ID NO 153; 3205pp; English.

XX The invention relates to novel genes ADA39629-ADA40565 and proteins

CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,

CC treating or ameliorating medical conditions e.g. by protein or gene

CC therapy. The polypeptides, nucleic acid molecules, antibodies or their

CC fragments, and agonists or antagonists that bind to the polypeptide are
CC useful for preparing a diagnostic or pharmaceutical composition for
CC diagnosing or treating cancer or other hyperproliferative disorder. The
CC polypeptides and nucleic acid molecules are also useful for detecting,
CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC or other hyperproliferative disorders including neoplasms, autoimmune
CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC thrombocytopenia), allergic reactions including asthma or eczema,
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC fungal or viral infections including HIV/AIDS), or wound healing and
CC disorders of epithelial cell proliferation. The nucleic acids are also
CC useful for chromosome identification, radiation hybrid mapping or long-
CC range restriction mapping, as molecular weight markers, or as
CC hybridization or diagnostic probes. The polypeptides and antibodies are
CC useful for providing immunological probes for differential identification
CC of the tissues immunohistochemistry assays. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pot_sequences.

SQ Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;

Query Match 99.9%; Score 1432; DB 7; Length 1434;
Best Local Similarity 100.0%; Pred. No. 1.5e-225;
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATTAACTCTTTTATCGGAATAGTATGATATTTCAATGTCACCTCCATTCAATGTTGA 60
DB 1 CATTAACTCTTTTATCGGAATAGTATGATATTTCAATGTCACCTCCATTCAATGTTGA 60
QY 61 TTTGGAGCTGACAGTATTTTGTGAAGCAGAGATTTAAATTTATATTGAAAGTCAGTGC 120
DB 61 TTTGGAGCTGACAGTATTTTGTGAAGCAGAGATTTAAATTTATATTGAAAGTCAGTGC 120
QY 121 AAAATTATGAATAGGATATCTAATAATCAAAAGTAAATCAAAAGTCAGAGTGT 180
DB 121 AAAATTATGAATAGGATATCTAATAATCAAAAGTAAATCAAAAGTCAGAGTGT 180
QY 181 CTAATATAAAATCTCGGTTCTTAAATTTATTTAAATTTATCTTGAATAGTTTCT 240
DB 181 CTAATATAAAATCTCGGTTCTTAAATTTATTTAAATTTATCTTGAATAGTTTCT 240
QY 241 TAGATTAACTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCAATAAA 300
DB 241 TAGATTAACTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCAATAAA 300
QY 301 CAAATTTCTATTAATGCMAGAGTGATATATACAGAAATTTATCAGGCATTACCAAGTC 360
DB 301 CAAATTTCTATTAATGCMAGAGTGATATATACAGAAATTTATCAGGCATTACCAAGTC 360
QY 361 TAGGCACATATAGGAATGACGACTCAGAAATGTTTCAATGTAGTATGATCTTGTA 420
DB 361 TAGGCACATATAGGAATGACGACTCAGAAATGTTTCAATGTAGTATGATCTTGTA 420
QY 421 AGGTAGGGAGCTTATTCAGACATAGTAGATAGTTTCTCTAAATGCTGTCTCAATGCTGG 480
DB 421 AGGTAGGGAGCTTATTCAGACATAGTAGATAGTTTCTCTAAATGCTGTCTCAATGCTGG 480
QY 481 CCTTTGGCTACTCTTCTCCATTTATGCGAGCCCATTCAGTCTTCAAGTTTCTTCT 540
DB 481 CCTTTGGCTACTCTTCTCCATTTATGCGAGCCCATTCAGTCTTCAAGTTTCTTCT 540
QY 541 GGACACCTTATGCTCTGAAATCATGACGAGGCTGATTCAATTTGGTGAATGGGTAGAAA 600
DB 541 GGACACCTTATGCTCTGAAATCATGACGAGGCTGATTCAATTTGGTGAATGGGTAGAAA 600
QY 601 GCAGTATGTTTGTGACATTAAGATGATAGTTTATAGATAGTTTATGCTTTAAGTGTAT 660
DB 601 GCAGTATGTTTGTGACATTAAGATGATAGTTTATAGATAGTTTATGCTTTAAGTGTAT 660

DB 601 GCAGTATGTTTGTGACATTAAGATGATAGTTTATAGATAGTTTATGCTTTAAGTGTAT 660
QY 661 GTTTTATATCTTTAAATAAAGAAATTAACCTTTTAAAGTATTCACCTCTCCCTCCAGC 720
DB 661 GTTTTATATCTTTAAATAAAGAAATTAACCTTTTAAAGTATTCACCTCTCCCTCCAGC 720
QY 721 CTAATCAAACTCGTGGATATATGAGAGATCTTGAAGAAGTAAATAAACCCTTCACT 780
DB 721 CTAATCAAACTCGTGGATATATGAGAGATCTTGAAGAAGTAAATAAACCCTTCACT 780
QY 781 GCTCCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTAATAC 840
DB 781 GCTCCCACTCCAGGTGAATCCGCCACTCCCACTGACCTAGTAGAATTTGTAATTAATAC 840
QY 841 TTACCTTCTATTTCTGAAATCAGTTGTGAACCTTGTGCTTATGTTTCAGARGTTTAAAGAC 900
DB 841 TTACCTTCTATTTCTGAAATCAGTTGTGAACCTTGTGCTTATGTTTCAGARGTTTAAAGAC 900
QY 901 CTCMGTCGAATTCATTTTAAATCTGTAATCTGGAAGCATTTGAATGAATTTCTTAACA 960
DB 901 CTCMGTCGAATTCATTTTAAATCTGTAATCTGGAAGCATTTGAATGAATTTCTTAACA 960
QY 961 AGAAGACTCATCTGTAGTCTTTTGTGACTCTCTATGAGCCCATTAAGGGTTCTGTGCTTA 1020
DB 961 AGAAGACTCATCTGTAGTCTTTTGTGACTCTCTATGAGCCCATTAAGGGTTCTGTGCTTA 1020
QY 1021 GCATTAACAAAATAAGSTTTATAGGTAAGCCCAATGTAATTTTGTGATGAGS 1080
DB 1021 GCATTAACAAAATAAGSTTTATAGGTAAGCCCAATGTAATTTTGTGATGAGS 1080
QY 1081 GCTTTAAATTTGTGCTCTTTTCAATTTTATTTTATTTTCAATTTATGTTTGTAACTG 1140
DB 1081 GCTTTAAATTTGTGCTCTTTTCAATTTTATTTTATTTTCAATTTATGTTTGTAACTG 1140
QY 1141 CTTTGTAGGAGATTAATATATGTTTAAATTTAGTTTGTGGGGGAAATAATTTGTCAAAGA 1200
DB 1141 CTTTGTAGGAGATTAATATATGTTTAAATTTAGTTTGTGGGGGAAATAATTTGTCAAAGA 1200
QY 1201 GGATTAATTTAATTTAGTGTCTTCTGTTTATTTTCAAGTAAGAGAGAGACCTACGCTGATA 1260
DB 1201 GGATTAATTTAATTTAGTGTCTTCTGTTTATTTTCAAGTAAGAGAGAGACCTACGCTGATA 1260
QY 1261 TTCAAGAGTTGTACCTTAACATTTGTTGAAACATTTTCTTAAAGATTTTCAAAAGGAATAT 1320
DB 1261 TTCAAGAGTTGTACCTTAACATTTGTTGAAACATTTTCTTAAAGATTTTCAAAAGGAATAT 1320
QY 1321 GTGTAAATTCAGAAATCATACCACTGTCTTAACTTGGTAAACAAACTGTTCTTAAATAA 1380
DB 1321 GTGTAAATTCAGAAATCATACCACTGTCTTAACTTGGTAAACAAACTGTTCTTAAATAA 1380
QY 1381 AGTATTTAATGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1434
DB 1381 AGTATTTAATGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1434

RESULT 2
ACCS0424

ID ACCS0424 standard; cdna; 1434 BP.

XX ACCS0424;

XX AC

XX AC

XX 12-JUN-2003 (first entry)

XX Human secreted protein coding sequence, SEQ ID 91.

XX Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
XX vulnery; antiinflammatory; nootropic; neuroprotective;
XX antiparkinsonian; gene therapy; human; cardiovascular disorder; gene; ss.

XX Homo sapiens.

XX OS

XX PN

XX WO200295010-A2

XX

RESULT 3

AA37452
ID AAX37452 standard; cDNA; 1434 BP.XX AC
XX AAX37452;

XX 06-JUL-1999 (first entry)

XX DE Human secreted protein cDNA fragment containing gene 2.

XX Human; secreted protein; treatment; prevention; protein therapy; AIDS;
XX gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
XX developmental abnormality; fetal deficiency; blood disorder; leukemia;
XX immune system disease; autoimmune disease; hepatic disease; lymphoma;
XX renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
XX cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
XX pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
XX arthritis; malignancy; digestive; endocrine; infection; ss.

XX OS Homo sapiens.

XX XX

XX PN W0918208-A1.

XX XX
XX PD 15-APR-1999.

XX XX 01-OCT-1998; 98WO-US020775.

XX PR 02-OCT-1997; 97US-0060833P.

XX PR 02-OCT-1997; 97US-0060836P.

XX PR 02-OCT-1997; 97US-0060837P.

XX PR 02-OCT-1997; 97US-0060838P.

XX PR 02-OCT-1997; 97US-0060839P.

XX PR 02-OCT-1997; 97US-0060843P.

XX PR 02-OCT-1997; 97US-0060862P.

XX PR 02-OCT-1997; 97US-0060866P.

XX PR 02-OCT-1997; 97US-0060874P.

XX PR 02-OCT-1997; 97US-0060880P.

XX PR 02-OCT-1997; 97US-0060884P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX XX

XX Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;
XX Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Peng P;
XX Lafleur DW, Shi Y;

XX WFL; 1999-264022/22.

XX P-PSDB; AAY07853.

XX PT New isolated human genes and the secreted polypeptides they encode.
XX Claim 1a; Page 228; 368pp; English.

XX This invention describes novel isolated human genes and the secreted
XX proteins they encode. The products of the invention are useful for
XX preventing, treating or ameliorating medical conditions, e.g. by protein
XX or gene therapy. Also pathological conditions can be diagnosed by
XX determining the amount of the new polypeptides in a sample or by
XX determining the presence of mutations in the new polynucleotides.

XX Specific uses are described for each of the 101 polynucleotides, based on
XX which tissues they are most highly expressed in, and include developing
XX products for the diagnosis or treatment of cancer, tumours,
XX neurodegenerative disorders, developmental abnormalities and fetal
XX deficiencies, blood disorders, leukemias, diseases of the immune system,
XX autoimmune diseases, hepatic and renal diseases, lymphomas, inflammation,
XX allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
XX disease, skeletal or cardiac muscle disorders, pulmonary disorders,
XX transplant rejection, disorders involving osteoclasts such as
XX osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
XX infections and AIDS. The human secreted proteins of the invention are
XX represented in AAY07852-Y07993 and the encoding nucleic acids are
XX represented in AAX37451-X37552

SQ Sequence 1434 BP; 480 A; 204 C; 250 G; 495 T; 0 U; 5 Other;

Query Match 99.7%; Score 1430.4; DB 2; Length 1434;
Best Local Similarity 99.9%; Pred. No. 2.7e-225;
Matches 1433; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTGCACATCCATTCATGTTGA	60
DB	1	CATTAAACTCTTTTATCGGGAATAGTATGATATTTTCAATGTGCACATCCATTCATGTTGA	60
QY	61	TTTGGAGCTGACAGTATTATTTTGTGTAAGCAGAGATTTAAATTTTATATTTGAAGTCAGTGC	120
DB	61	TTTGGAGCTGACAGTATTATTTTGTGTAAGCAGAGATTTAAATTTTATATTTGAAGTCAGTGC	120
QY	121	AAAATTATGATAGGATATCTAATAATAACAAGTAAATAACAAGTCAAGCAGGTTT	180
DB	121	AAAATTATGATAGGATATCTAATAATAACAAGTAAATAACAAGTCAAGCAGGTTT	180
QY	181	CTAAATAAAAATTTCTGGGTTCTTAAATAATTTTAAATTTATCTTGAATAGTTTCT	240
DB	181	CTAAATAAAAATTTCTGGGTTCTTAAATAATTTTAAATTTATCTTGAATAGTTTCT	240
QY	241	TAGATTAACTCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATPAAA	300
DB	241	TAGATTAACTCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATPAAA	300
QY	301	CAAAATTGTCTATTAAATGCMAGAGTGGTAATATACAGAAATTTATCAGGCATTTACCAAGTC	360
DB	301	CAAAATTGTCTATTAAATGCMAGAGTGGTAATATACAGAAATTTATCAGGCATTTACCAAGTC	360
QY	361	TAGGCACATATAGGAATGACAGCACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTGTA	420
DB	361	TAGGCACATATAGGAATGACAGCACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTGTA	420
QY	421	AGGTAGGGAGCTTATTACAGACATAGTAGATAGTTTCTCTAAATGCTGTSTCAATTTGCTG	480
DB	421	AGGTAGGGAGCTTATTACAGACATAGTAGATAGTTTCTCTAAATGCTGTSTCAATTTGCTG	480
QY	481	CTTTGGCTACCTGTACTTCCSCAATATGCGACGCCATTCAGTCTTGAGTTTCTTCTCT	540
DB	481	CTTTGGCTACCTGTACTTCCSCAATATGCGACGCCATTCAGTCTTGAGTTTCTTCTCT	540
QY	541	GGACACCTTATGCTCTGAAATCATGACGAGGCTGATTCATTTGGTGATTTGGGTAGAAA	600
DB	541	GGACACCTTATGCTCTGAAATCATGACGAGGCTGATTCATTTGGTGATTTGGGTAGAAA	600
QY	601	GCAGTATGTTTTGCTGACATTAAGATGTAGGTATAGATAGTGTAGCTTTAAGTGTAT	660
DB	601	GCAGTATGTTTTGCTGACATTAAGATGTAGGTATAGATAGTGTAGCTTTAAGTGTAT	660
QY	661	GTTTTTATCTTTTAAATAAGAAATATACCTTTTAAAGCTATTTCACCTCTCTCCCCAGC	720
DB	661	GTTTTTATCTTTTAAATAAGAAATATACCTTTTAAAGCTATTTCACCTCTCTCCCCAGC	720
QY	721	CTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAATCACTTCACT	780
DB	721	CTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAATCACTTCACT	780
QY	781	GCTCCACTCCAGTGAATCGCCCACTCCCACTGACCTAGTAGAAATTTGTAATTTAATAC	840
DB	781	GCTCCACTCCAGTGAATCGCCCACTCCCACTGACCTAGTAGAAATTTGTAATTTAATAC	840
QY	841	TTACCTTCTATTCTCAAAATCAGTGTGAACTGTGCTTATGCTTCTCAGAGTTTAAAGAAC	900
DB	841	TTACCTTCTATTCTCAAAATCAGTGTGAACTGTGCTTATGCTTCTCAGAGTTTAAAGAAC	900
QY	901	CTCMGTGAATTCATTTTTTAAATCTGCTATTCTGAGAAGCAATTTGAATGAATTTCTTAACA	960
DB	901	CTCMGTGAATTCATTTTTTAAATCTGCTATTCTGAGAAGCAATTTGAATGAATTTCTTAACA	960
QY	961	AGGAGACTCATCTGAGCTGTTTGTGCTGACTCTCTAGGCCCATAGGGTCTGTGCTTA	1020
DB	961	AGGAGACTCATCTGAGCTGTTTGTGCTGACTCTCTAGGCCCATAGGGTCTGTGCTTA	1020


```

PR 27-OCT-1998; 98US-0105877P.
XX (CHIR ) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI Lamson G, Drmanac R, Kravjenjakov R, Dickson M, Drmanac S, Labat I;
PI Leehkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-293155/25.
XX Polynucleotide library comprising 1079 defined sequences, useful in the
PT form of an array to detect cancer or susceptibility to cancer.
XX Claim 1; Page 268; 502pp; English.
XX The present invention describes a library of polynucleotides comprising
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC one of the 1079 sequences; (2) a recombinant host cell containing (I);
CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
CC of detecting differentially expressed genes correlated with a cancerous
CC state of a mammalian cell comprising detecting a gene product encoded by
CC 65 of the 1079 sequences given in the specification. The polynucleotides
CC are used to monitor patients having (or susceptible) to cancer to detect
CC potentially malignant events at a molecular level before they are
CC detectable at a gross morphological level. The polynucleotides are also
CC useful for monitoring the efficacy of various therapies and preventive
CC interventions. Polynucleotide probes based on the disclosed sequences are
CC useful for chromosome mapping and detection of transcription levels. The
CC 1079 polynucleotide sequences were derived from a human colon cancer cell
CC line Km12L4-A cDNA library
XX Sequence 415 BP; 114 A; 73 C; 87 G; 141 T; 0 U; 0 Other;
XX Query Match 25.1%; Score 360; DB 3; Length 415;
XX Best Local Similarity 94.4%; Pred. No. 4e-50;
XX Matches 391; Conservative 3; Mismatches 18; Indels 2; Gaps 2;
QY 302 AAATTGCTATTAAATGAMGA-GTGTAATATACAGAAATTAATCAGGCAATACCAAGTC 360
DB 1 AAATTGCTATTAAATGAMGA-GTGTAATATACAGAAATTAATCAGGCAATACCAAGTC 60
QY 361 TAGGCACATATAGGAATGAGCACTCAGAAATGTTTCAATGTAGTGTGATGCTGTGA 420
DB 61 TAGGCACATATAGGAATGAGCACTCAGAAATGTTTCAATGTAGTGTGATGCTGTGA 120
QY 421 AGGTAGGGAGCTTATTTCAGACATAGTAGATAGTTTCTTAATGCTGTSTCAATTTGCTGG 480
DB 121 AGGTAGGGAGCTTATTTCAGACATAGTAGATAGTTTCTTAATGCTGTSTCAATTTGCTGG 180
QY 481 CCTTTGGCTACCTGTACTTCGCAATATGAGGAGCCCAATTCAGTCTTGAGTTTCTCTCT 540
DB 181 CCTTTGGCTACCTGTACTTCGCAATATGAGGAGCCCAATTTGCGGCTT-TTTGTTCTCTCTG 239
QY 541 GGCACACTTATGCTCTGAATCATGAGCGAGGCTGATTCAATTTGGTATTGGTAGAAA 600
DB 240 GGCACACTTATGCTCTGAATCATGAGCGAGGCTGATTCAATTTGGTATTGGTAGAAA 299
QY 601 GCAGTATGTTTCTGCAATTAAGATAGTATAGTTATAGATAGTTTACGCTTTTAAGTGTAT 660
DB 300 GCAGTATGTTTCTGCAATTAAGATAGTATAGTATAGTATAGTATAGTATAGTATAGTAT 359
QY 661 GTTTTTATACCTTTAAATAAGAAATATAACCTTTTAAAGCTTATTCACCTCTCC 714
DB 360 GTTTTTATACCTTTAAATAAGAAATATAACCTTTTAAAGCTTATTCACCTCTCC 413
XX RESULT 6
XX AAC10084
XX ID AAC10084 standard; cDNA; 395 BP.
XX AAC10084;
XX 06-OCT-2000 (first entry)
XX Human secreted protein 5' EST, SEQ ID NO: 14159.
XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EPI033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-00200610.
XX 26-FEB-1999; 99US-0122487P.
XX (GEST ) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 1; SEQ ID NO 14159; 71pp + Sequence Listing; English.
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
XX ESTs are derived from mRNAs with intact 5' ends and can therefore be used
XX to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors
XX Sequence 395 BP; 147 A; 41 C; 74 G; 125 T; 0 U; 8 Other;
XX Query Match 24.9%; Score 356.8; DB 3; Length 395;
XX Best Local Similarity 95.7%; Pred. No. 1.3e-49;
XX Matches 378; Conservative 8; Mismatches 6; Indels 3; Gaps 2;
QY 58 TGATTTGGAGCTGACAGTATTATTTTGTGTAGCAGAGATTTAATTTATATTTGAAGTCAG 117
DB 1 TGATTTGGAGCTGACAGTATTATTTGTGTAGCAGAGATTTAATTTATATTTGAAGTCAG 60
QY 118 TGCATAATTTAGTATAGGATATATTAATAACAAGTAATAACAAGTCAAGACAGT 177
DB 61 TGCATAATTTAGTATAGGATATATTAATAACAAGTAATAACAAGTCAAGACAGT 120
QY 178 GTTCTAAATAAATAATCTGGTTCCTTAAATAATTTTAAATTT--ATCTTCAAAATAGT 235
DB 121 GTTCTAAATAAATAATCTGGTTCCTTAAATAATTTTAAATTTTATCTTTGAAATAGT 180
QY 236 TTTCTTAGATTAATCTCAGATATGAGAAAGTCAATTAAGTGTGAGTAAGTAGTATCA 295
DB 181 NMTCTTAGATTAATCTCAGATATGAGAAAGTCAATTAAGTGTGAGTAAGTAGTATCA 240
QY 296 TTAACAAATTTGCTATTAAATGCMGA-GTGTAATATACAGAAATTTATCAGGCATTAC 354
DB 241 TTAACAAATTTGCTATTAAATGCMGA-GTGTAATATACAGAAATTTATCAGGCATTAC 300

```

D	b	61	CAC	TGACCTAGT	AGAATTTGTAAATTTAATAACTTTACCTTCTATTTCGAAAATCAGTTGTGA	120
Q	y	870	ACT	GTTGCCCTTA	TGCTTCAGARGTTTAAAGAACCTCMGTGAATCATTTTTTAAAAATCTGCT	929
D	b	121	AC	TGTTGCCCTTA	TGTTTGAGAGTTTAAAGAACCTCAGTGAATTCATTTTTTAAAAATCTGCT	180
Q	y	930	ATT	CTGAGAAGCA	TGTAATGAATCTTTTAAACAAGAAGACTCATCTGTAGCTGTTTGTCTGAC	989
D	b	181	ATT	CTGAGAAGCA	TGTAATGAATCTTTTAAACAAGAAGNCTCATCTGTAGCTGTTTGTCTGAC	240
Q	y	990	TC	TATGAGCCCA	TAAAGGGTCTGTGCCTTAGCATTTAAACAATAAGCTTTATAGG--TA	1047
D	b	241	TC	TATGAG--	CCCATAAAGGGTCTGTGCCTTAGCATTTAAACAATAAGGTTTATAGGGTAA	299
Q	y	1048	AAG	CAATGTATTA	ATATTTTTTTTGGCATGGAGGC	1082
D	b	300	AAG	CAATGTATTA	NTTTTTTTTTTGGCATGGGGGC	334
R	E	S	U	L	T	
I	D	A	H	A	H	70126/c
I	D	A	H	A	H	70126 standard; cDNA; 545 BP.
X	X	A	A	H	70126;	
X	X	19-SEP-2001	(first entry)			
X	X	Human cervical cancer marker nucleic acid	1400.			
X	X	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.				
X	X	Homo sapiens.				
X	X	W0200142467-A2.				
X	X	14-JUN-2001.				
X	X	08-DEC-2000; 2000WO-US033312.				
X	X	08-DEC-1999; 99US-0169681P.				
X	X	21-DEC-1999; 99US-0171350P.				
X	X	14-MAR-2000; 2000US-0189315P.				
X	X	12-MAY-2000; 2000US-0203791P.				
X	X	09-JUN-2000; 2000US-0210600P.				
X	X	21-JUL-2000; 2000US-0220114P.				
X	X	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.				
X	X	Schlegel R, Deeds J, Berger A, Zhao X;				
X	X	WPI; 2001-375006/39.				
X	X	New isolated nucleic acid for diagnosing and treating cervical cancer and for assessing and detecting compounds for treating the cancer.				
X	X	Claim 1; Page 319-320; 1051pp; English.				
X	X	The invention relates to novel genes (AAH68727-AAH73383) associated with cervical cancer with cytostatic activity. The nucleic acids and encoded polypeptides are useful: to assess if a patient is afflicted with cervical cancer or has a pre-malignant condition; to monitor the progression of cervical cancer or a premalignant condition in a patient; and to select and/or assess the efficacy of a compound or therapy for inhibiting cervical cancer in a patient. The nucleic acids may also be useful for gene therapy				
X	X	Sequence 545 BP; 200 A; 40 C; 23 G; 209 T; 0 U; 73 Other;				
Q	Y	Query Match	4.9%;	Score 69.8;	DB 4;	Length 545;
Q	Y	Best Local Similarity	43.9%;	Pred. No. 0.013;	Mismatches 167;	Conservative 0;
Q	Y	Matches 167;	Conservative 0;	Mismatches 213;	Indels 0;	Gaps 0;
Q	Y	1055 TG	TATTAATATTTTTTTTTTG	CATGAGCGCTTAAAAATTTG	TCTCTCTTTTCATATTTAT	1114

Db 410 TTTTAAANNTTTNNNNNTTTTAAAAAANNTTTTNTTNAATTTTTTTA 351
 QY 1115 CATATTCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATATATGTTATAAATTAG 1174
 Db 350 AATTGGAATTTTAAAAANTANTGNTNAACGNGNCRAAAAATNAANTTTTTNTTNNNT 291
 QY 1175 TTTTGGGGGAATAATGTGCAAGAGATAATTTAATTTACGTGCTTCTGTTATTCAGA 1234
 Db 290 TTTTNAAAAAANNTTTTNCCTNNNANTTTNNNATNTTTTTTTTTTGNANTTTTTTCCC 231
 QY 1235 ATAAAGAGAGACGACGCTGCATATTCAGAGTTGCTACCTTAACATTCGTGAACATT 1294
 Db 230 CCCAANTTTTTTTNNNNNTTTTTTTNNNNNTTTTNNNTTTTNNNTTTTNNAAAAAANTTT 171
 QY 1295 TTTTCTAAGATTTTCAAAAGGAATATGTGTAAATTTGAGAAATCATACCACTGTCCTAAC 1354
 Db 170 TTTTNAANTTTTTTNAAAAAAANNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 111
 QY 1355 TTGGTAAACAACTGTTCTTAATAAGTATTTAATGATTTTAAAAAANNTTTTTAAAAA 1414
 Db 110 AAAAAAANNTTT 51
 QY 1415 AAAAAAANNTTT 1434
 Db 50 GAAAAAANNTTT 31

RESULT 9
 ABZ10246/c
 ID ABZ10246 standard; DNA; 8056 BP.
 XX
 AC ABZ10246;
 XX
 DT 16-JAN-2003 (first entry)
 XX
 DE Haematopoietic cell proliferation disorder related DNA sequence #386.
 XX
 KW Human, haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO20027272-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-EP003401.
 XX
 PR 26-MAR-2001; 2001US-0278333P.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
 PI Schwöpe I, Ziebarth H;
 XX
 DR WPI; 2003-018942/01.
 XX
 PT Detecting and differentiating between hematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent that
 PT distinguishes between methylated and non-methylated CpG dinucleotides.
 XX
 PS Claim 28; SEQ ID NO 386; 117pp; English.
 XX
 CC The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG

CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclases, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 SQ Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 U; 0 Other;
 Query Match 4.8%; Score 69; DB 7; Length 8056;
 Best Local Similarity 48.0%; Pred. No. 0.017;
 Matches 293; Conservative 1; Mismatches 306; Indels 11; Gaps 3;
 QY 826 TTGTAAATTAATACCTTCTCTATTCTGAATCAGTTGTGAACCTGTTCCTATGTT 885
 Db 2006 TTATTAAATTTTAAATAAATAAATTAATTAATTAATTTTATTAATAATTTATTTT 1947
 QY 886 CAGAGTTTAAAGAACCTCMGTGAATTCATTTTTTAAATCGTATTTCTGAGAGCATTTG 945
 Db 1946 TATTATATTAATTAATTAATTTTATTTTATTAATAATTTTATTTATTAATAAAA 1887
 QY 946 AATGAATCTTACAGAGAGACTCATCTGTAGCTGTTGTGCTCCTATGAGCCCAT 1005
 Db 1886 AATAAATAAATAAATAAATAAATTTTATTTTATTTTATTTTATTAACAAAA 1827
 QY 1006 AGGGTCTGTGCTTAGCAATTA---ACAAATAAGGTTTAGTGAAGCAATGATTA 1061
 Db 1826 TTTTATTATTATTAAATTAATAATAATAATAATAATAATAATAATAATAATAATA 1767
 QY 1062 ATTTTTTTTGTGATGAGGCTTTAAATTTGCTGCTTTTTCATATTTTATTCATATTC 1121
 Db 1766 AAAAAATTTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTATATATTT 1707
 QY 1122 AATTTATGTTTGTAACTGCTTTTATAGGAGATAATTAATTTATATATTTAGTTTGG 1181
 Db 1706 AAATTAATTTTATTAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAA 1649
 QY 1182 GGAATAAATTTGTCAAGAGGATAATTTAATTTACGTGCTTCTGTTATTCAGATAAAGA 1241
 Db 1648 TTTTTTAATTTTACAA---TTCAATTTATTTCAATTTTAAAAATATTAATAA 1594
 QY 1242 GAGAAGACTACGCTGCATATTCAGAGTTGTACCTTAAACATTCGTGTAACATTTTCTA 1301
 Db 1593 TATAAATTAATAATTAATAATTTAAATTTTAAATAAATAAATAAATAAATAAATA 1534
 QY 1302 AGATTTTCAAAAGGATATGTTGAATTTGAGAATCATACCACTGTCCTAATCTGTAA 1361
 Db 1533 TTTTAAACATATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 1474
 QY 1362 ACAAACTGTTCTTTAAATAAAGTATTTTAAGTATTTAAAAAANNTTTTTAAAAA 1421
 Db 1473 ATATAATTTTATTAATAAATAAATAATTTTATTTTATTTTATTTTATTTTATTTA 1414
 QY 1422 AAAAAAANNT 1432
 Db 1413 TAAAAAANNT 1403
 RESULT 10
 ABX46069/c
 ID ABX46069 standard; cDNA; 419 BP.
 XX
 AC ABX46069;

Query Match	4.6%; Score 65.6; DB 4; Length 597;	
Best Local Similarity	49.7%; Pred. No. 0.061;	
Matches 196; Conservative	0; Mismatches 194; Indels 4; Gaps 1;	
1038	TTTATAGTAAAGCAATGATTAATTTTTTTTTTTCATGAGGGCTTTAAATTTCTGCT	1097
395	TTTTTAAAAAATAATAGTTTTTTTTTTTTTAAAGAACCTCTCAATTTTGGAA	336
1098	CTTTTTCATATTTTATTCATATTCAAATTTATGGTTTCTAACTGCTTTTAGGGAGATAAT	1157
335	ATTTTT---TTTTTTTTTTTTTCAAAAAATATTTTTTTTTTTTTTTTTTTTTTAAAAAAT	280
1158	TATATGTATATAATAGTTTTTGGGGGAATAATTTGTCGAAGAGGATAATTAATTTACG	1217
279	TTTTTTTTTTAAAAAACCCTGGGGTTTTTTTTTTTTTTTGAAGAAATTTTTTTTTTTT	220
1218	TGCTTCTGTTATTCAGAAATAAGAGAGAACTACGCTGCATATTCAGAGTTGTACCTT	1277
219	TTTCCCTCTCGGAATTTTAAAAAATTTTTTTTGTAAAAATTCGGGGGGTTTTT	160
1278	AACATTTGTGAACATTTTTTCTTAAGATTTTCAAAAGGAATATGTGTAAATTCAGAAATC	1337
159	TAAATTTGGGGATTTTTTTTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA	100
1338	ATAACCACTGCTCAACTTGCTGTAACAACTGTTCTTTAAATAAGTATTTAATGATTTTA	1397
99	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA	40
1398	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA	1431
39	AAAGAAGAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA	6
RESULT 12		
ABZ10246		
ID	ABZ10246 standard; DNA; 8056 BP.	
XX		
ABZ10246;		
16-JAN-2003	(first entry)	
XX		
XX	Haematopoietic cell proliferation disorder related DNA sequence #386.	
DE	Human; haematopoietic cell proliferation disorder; cytostatic;	
XX	gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;	
KW	cytosine methylation state; gene; ds.	
KW		
XX		
OS	Homo sapiens.	
WO20027272-A2.		
PN		
EPD	03-OCT-2002.	
XX		
XX	26-MAR-2002; 2002WO-EP003401.	
PF		
XX		
XX	26-MAR-2001; 2001US-0278333P.	
PR		
XX	(EPIG-) EPIGENOMICS AG.	
XX	Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;	
PPI	Olek A, Piepenbrock C, Adorjan P, Grabs G, Iesche R, Leu E;	
PPI	Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;	
PPI	Schwöbe I, Ziebarth H;	
XX		
XX	WPI; 2003-018942/01.	
DR		
XX	Detecting and differentiating between hematopoietic cell proliferative	
XX	disorders, comprises contacting a target nucleic acid with a reagent that	
PPT	distinguishes between methylated and non-methylated CpG dinucleotides.	
PPT		
XX		
XX	Claim 28; SEQ ID NO 386; 117pp; English.	
PP		
XX	The present invention describes a method for detecting and	
CC		

ABZ10100/c
 ID ABZ10100 standard; DNA; 8056 BP.
 XX
 AC ABZ10100;
 XX
 DT 16-JAN-2003 (first entry)
 XX
 DE Haematopoietic cell proliferation disorder related DNA sequence #240.
 XX
 KW Human; haematopoietic cell proliferation disorder; cytostatic;
 KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KW cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FN W0200277272-A2.
 XX
 PD 03-OCT-2002.
 XX
 XX 26-MAR-2002; 2002WO-EP003401.
 PF
 XX 26-MAR-2001; 2001US-0278333P.
 PR
 XX (BPIG-) EPIGENOMICS AG.
 PA
 XX Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorian P, Grabs G, Lesche R, Leu E;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T, Pelet C;
 PI Schwöbe I, Ziebarth H;
 XX WPI; 2003-018942/01.
 DR
 XX
 XX
 XX
 PT Detecting and differentiating between hematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent that
 PT distinguishes between methylated and non-methylated CpG dinucleotides.
 XX
 XX Claim 28; SEQ ID NO 240; 117pp; English.
 PS
 XX
 XX
 CC The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related DNA
 CC sequences. The nucleotide sequences from the present invention can also
 CC be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables a
 CC highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients
 XX
 SQ Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 U; 0 Other;
 Query Match 4.4%; Score 62.6; DB 7; Length 8056;
 Best Local Similarity 47.3%; Pred. No. 0.19;
 Matches 289; Conservative 1; Mismatches 310; Indels 11; Gaps 3;
 QY 826 TTGTAAATTAATCTTACCTCTATTCTGAATCAGTTGCACTGTTGCCTTAGTT 885
 Db
 2006 TTATTAATTTTAAATAAAAAATTAATTAATTAATTTTAAATTAATTTATTTT 1947
 QY 886 CAGAGTTTAAAGACCTCMGTGATTCATTTTAAATCTGCTATTCAGAGCATTC 945

Db 1946 TATTTATATATAAATTAATTAATTTTATTTTATATAAAATTTTATTTTATATAAAAA 1887
 QY 946 AATGAATTCCTTAACAAGAGACTCATCTGTAGCTGTTGCTGACTCTCTATGAGCCCATTA 1005
 Db 1886 AAATAAAAAATAATTAATTAATTAATTTTATTTTATTTTATTTTATTTTATTTTAAACGAAA 1827
 QY 1006 AGGGTTCGTGCTTAGCATTA---ACAAAATAAGGTTTATAGTAAAGCCAATGATTA 1061
 Db 1826 TTTTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTTTATTTTATTTTATTTA 1767
 QY 1062 ATTTTATTTTGCATGGAGGCTTTAAAAATTTGTGCTCTTTTTCATATTTTATTCATATTC 1121
 Db 1766 AAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1707
 QY 1122 AATTATGTTTGTAACTGCTTTTATAGGAGATAATTAATTAATTAATTAATTAATTAATTTGGG 1181
 Db 1706 AAATTAATTTTATATAAT--TTAATAAAAAATAAAATTTTATCGTATTAATTAATAAATAAT 1649
 QY 1182 GCGAATAATTTGTCAAAAGAGGATAATTAATTTACGTGCTTCTGTTATTCAGATAAAGA 1241
 Db 1648 TTTTATTAATTTTACGATTCGTTTATTTTCGATTT-----TTTAAATATTAATAATTAATA 1594
 QY 1242 GAGAAGACTACGCTGCATATTCAGAGTTGTACCTTTAACAATTCGTAACCTGCTTAACTTGGTAA 1301
 Db 1593 TATAAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1534
 QY 1302 AGATTTTCAAAAGGATAATGTTAAATTTGAGAATTCATAACCACTGCTTAACTTGGTAA 1361
 Db 1533 TTTTATTAACGTATTTATATTTTATATTTATATATTAATTAATTAATTAATTAATTAATTTTA 1474
 QY 1362 ACAAACTGTTCTTAAATAAAGTATTTAATGATTTTAAATAAAAAATTTTATTTTATTTTATTT 1421
 Db 1473 ATATAATTTTATTAATAAATAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1414
 QY 1422 AAAAAAATAA 1432
 Db 1413 TAAAAAATAA 1403
 RESULT 15
 AAD08708
 ID AAD08708 standard; cDNA; 2103 BP.
 XX
 AC AAD08708;
 XX
 DT 04-SEP-2001 (first entry)
 XX
 DE Lycopersicon pennellii elin7 invertase isoenzyme cDNA.
 XX
 KW Invertase isoenzyme; apoplast; brix value; monosaccharide; sugar; fruit;
 KW green-fruited tomato; elin7; ss.
 XX
 OS Lycopersicon pennellii.
 XX
 FH Key Location/Qualifiers
 CDS 82..1833
 FT /*tag= a
 FT /product= "elin7 invertase isoenzyme"
 XX
 PN W0200149826-A1.
 XX
 PD 12-JUL-2001.
 XX
 PF 02-JAN-2001; 2001WO-IL000008.
 XX
 PR 04-JAN-2000; 2000US-00477375.
 XX
 PA (YISS) YISSUM RES DEV CO HEBREW UNIV JERUSALEM.
 XX
 PI Zamir D, Pleban T, Fridman E;
 XX WPI; 2001-418359/44.
 DR P-PSDB; AAE04416.
 DR

```
XX Nucleic acid comprising a genomic, complementary or composite sequence
PT encoding a polypeptide having an invertase activity in an apoplastic
PT environment, useful for increasing the monosaccharide content of plants.
XX
XX Claim 3; Page 94; 99pp; English.
XX
CC The present invention relates to nucleic acids comprising genomic,
CC complementary or composite polynucleotide sequences encoding novel
CC apoplastic invertase isoenzymes associated with high brix value in
CC Lycopersicon pennellii fruit, and an N terminal amino acid sequence
CC serving for secretion into an apoplast. The invention also relates to a
CC novel regulatory element unique to L. pennellii which is also associated
CC with high brix trait. The polynucleotides of the invention are useful for
CC increasing the sugar content in e.g. fruits, roots and leaves of plants
CC expressing them. The regulatory element when integrated, in a site
CC specific manner, into a solanaceae plant genome, can substantially
CC increase the sugar content in tissues e.g. fruits. The present sequence
CC is Lycopersicon pennellii cDNA encoding elin7 invertase isoenzyme
XX
SQ Sequence 2103 BP; 742 A; 353 C; 418 G; 590 T; 0 U; 0 Other;
Query Match 4.3%; Score 62; DB 5; Length 2103;
Best Local Similarity 53.3%; Pred. No. 0.24;
Matches 131; Conservative 0; Mismatches 115; Indels 0; Gaps 0;
QY 1189 ATTGTGCAAGAGGATAATTAAATTTACGTCTTCTGTATTTCAGAAATAAGAGAGAAGA 1248
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1249 CTACGCTGCATATTCAAGAGTTGTACCTTAAACATTGGTGAAACATTTTCTTAAGATTTT 1308
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1309 CAAAAGGAATATGTGTAATTTGAGAAATCATACCACTGTCCTAACTTGTGTAACAACT 1368
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1369 GTTCTTTAAATAAAGTATTTTAATGATTTTAAAAAATAAATAAATAAATAAATAAATAA 1428
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1429 AAAAAA 1434
Db |||||
QY 2069 AAAAAA 2074
```

Search completed: July 20, 2004, 17:19:03
Job time : 587 secs

This Page Blank (uspto)


```
QY      1418 AAAAAAAAAAAAAA 1432
          |::| | | | | |
Db      359 AWRACAAATATA 373
```

```

RESULT 2
US-09-621-976-16015
; Sequence 16015, Application US/09621976
; Patent NO. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 16015
; LENGTH: 339
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-16015

```

Query Match	4.1%	Score 58.2;	DB 4;	Length 339;
Best Local Similarity	64.8%;	Pred. No. 0.0023;		
Matches 81; Conservative	3;	Mismatches 41;	Indels 0;	Gaps 0;
QY 1310	AAAAGCAATATGTAATTGAGAAATCATAAACCACTGCTCTAACTTGGTAAACAACACTG	1369		
Db				
QY 197	ATAGAAAATAATTCGATAGAAAATAATAGAAATAGAAAAAATTTTAAAAAACCAACCA	256		
Db				
QY 1370	TTCTTAAATAAAGCTATTTTAATGATTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA	1429		
Db				
QY 257	AGCCTCATCTATGAAAAAATAAAWAAAAAATAAAAAAAAAAAAAAAAAAAAAA	316		
Db				
QY 1430	AAAAA 1434			
Db				
QY 317	AAAAA 321			
Db				

```

RESULT 3
US-09-621-976-14761
; Sequence 14761, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 14761
; LENGTH: 189
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-14761

```

	Query Match	4.0%;	Score 57.6;	DB 4;	Length 189;
	Best Local Similarity	57.3%;	Pred. No. 0.0027;		
	Matches 102;	Conservative	1;	Mismatches 75;	Indels 0; Gaps 0;
1257	CATATTCAAGAGTGTGACCTTAACATGTGTGGAACAATTTTCTAGATTTC	CAAAAGCA	1316		
12	CAGATATCAAGATAGTATGTATACACTGCCATGTGACATATACATATGTGTGTGA	71			
1317	APATGTGTTAAATTGAGAAATCAPAACCACTGTCTCAACTTGGTAAACAAACTGTTCTTAA	1376			

Db 72 ATGTTATTATTATTTAAAAAATAAACACTGCTTACTTGAATAAAAAAAAAA 131

Qy 1377 ATAAAGTATTAAATGATTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA 1434

Db 132 AAAR 189

RESULT 4
 US-08-983-502-14
 ; Sequence 14, Application US/08983502
 ; Patent No. 6399327
 ; GENERAL INFORMATION:
 ; APPLICANT: David WALLACH
 ; APPLICANT: Mark P. BOLDIN
 ; APPLICANT: Tanya M. GONCHAROV
 ; APPLICANT: Yuri V. GOITSEV
 ; TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF FAS RECEPTORS
 ; TITLE OF INVENTION: AND OTHER PROTEINS
 ; NUMBER OF SEQUENCES: 34
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSES: Browdy and Neimark
 ; STREET: 419 Seventh Street N.W., Ste. 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/983,502
 ; FILING DATE: 16-JAN-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US96/10521
 ; FILING DATE: 14-JUN-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: IL 114,615
 ; FILING DATE: 16-JUL-1995
 ; APPLICATION DATA:
 ; APPLICATION NUMBER: IL 114,986
 ; FILING DATE: 17-AUG-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: IL 115,319
 ; FILING DATE: 14-SEP-1995
 ; APPLICATION DATA:
 ; APPLICATION NUMBER: IL 116,588
 ; FILING DATE: 27-DEC-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: IL 117,932
 ; FILING DATE: 16-APR-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Browdy, Roger L.
 ; REGISTRATION NUMBER: 25,618
 ; REFERENCE/DOCKET NUMBER: WALLACH-19
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 628-5197
 ; TELEFAX: (202) 737-3528
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2887 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: CDNA
 US-08-983-502-14

Query Match 3.9%; Score 56.4; DB 4; Length 2887;
Best Local Similarity 64.8%; Pred. No. 0.0078;
Matches 84; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
1305 TTTTCAAAGAGATATGTGTAAATGAGAAATCATACCACCTGTCTCACTTGGTAAACA 1360


```
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 235..399
; US-09-621-976-2813

Query Match          3.9%; Score 55.8; DB 4; Length 832;
Best Local Similarity 11.2%; Pred. No. 0.0083;
Matches 39; Conservative 169; Mismatches 141; Indels 0; Gaps 0;

QY 1085 TAAATTTGCTCTTTTTCATATTTTATTCATATTCATATTTATGTTGTTGTAAGTCTTT 1144
Db 373 TTATATTTTGTYYTWKWTWYWTWTTWTRWKKKRWYWWKSYACASRYKRW 314

QY 1145 TTAGGAGATAATATATGTTATATAATGTTTGGGGGAATAATGTCGAAGAGAT 1204
Db 313 GWMYMWKMWSTWYCMYCMKCMYRGRCAMWTMARGRWSYANGKWSMSMCT 254

QY 1205 AATTAATTTACGTCTCTGTTATTCAGATAAAGAGAAGAACTACGTCGATATCA 1264
Db 253 RMYYKSGSTYWKCTCATCYWYWKYKMKMSKTCWSGRGGYMTSYSTRYSYMWAS 194

QY 1265 AGAGTTGTCCTTAAACATTTGTTGAACATTTTCTTAAGATTTTCAAGGAATATGTT 1324
Db 193 WMTYMCWGWGWSWSTYWMYMWGKKWRYATTWRRMMWMAATMMYMWMAWCMRGA 134

QY 1325 AATAGAGAATCATACCACTGCTCTAACTGTTAAACAACTGTTCTTTAAATAAGTA 1384
Db 133 AMYRRTMMWGYRYWYWKSYRTRCAWAWKTKRSYWCWKKRCMMWMAWY 74

QY 1385 TTTAATGATTTTAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1433
Db 73 KTWMMRACWKTRYRWWAWAWRWWMTMMMYWYWRAMKRRWWRKWR 25

RESULT 8
US-07-945-288-9
; Sequence 9, Application US/07945288
; Patent No. 5433948
; GENERAL INFORMATION:
; APPLICANT: Thomas, Wayne R.
; APPLICANT: Chua, Kaw-Yan
; TITLE OF INVENTION: CLONING AND SEQUENCING OF ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES (HOUSE DUST MITES)
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655

; APPLICATION NUMBER: US/07/945,288
; FILING DATE: 19920910
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 580,655
; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: P36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1172 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..738
; US-07-945-288-9

Query Match          3.9%; Score 55.8; DB 1; Length 1172;
Best Local Similarity 61.2%; Pred. No. 0.0086;
Matches 90; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1288 AACAATTTTCTTAAGATTTTCAAGAGGATATGTTAAATTCAGAAATCATACCACTG 1347
Db 1017 AAAGACAATTTCTATATGATTCCTCAATAATTTTAAATCAAAATTTTGAAGAATG 1076

QY 1348 TCCTAACTTGGTAAACAACTGTTCTTAAATAAGTATTTAAATGATTTTAAAAAATA 1407
Db 1077 AATAAATTCATTCACAAATTTAAAAAATAAAAAAATAAAAAAATAAAAAAATA 1434

QY 1408 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1163
Db 1137 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1136

RESULT 9
US-08-462-831-9
; Sequence 9, Application US/08462831
; Patent No. 5552142
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,831
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655
```

```

; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: US 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1172 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..738
; US-08-462-831-9

Query Match 3.9%; Score 55.8; DB 1; Length 1172;
Best Local Similarity 61.2%; Pred. No. 0.0088;
Matches 90; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1288 AACATTTTCTTAAGATTTTCAAAAGGAATATGTGTAATTCAGAAATCATAACCACTG 1347
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1017 AAAGACAATTTCTATATGATGTCACATAATTTATTTAAATCAAAATTTTAGAAATG 1076
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1348 TCCTACTTGGTAAACAAACTGTTCTTAATAAAGTATTTAATGATTTTAAATGATTTTAA 1407
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1077 AATAAATTCATTCACAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTT 1136
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1408 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1137 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1163
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 10
US-08-461-809-9
; Sequence 9, Application US/08461809
; Patent No. 5770202
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,809
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655
; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: US 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
```

```

; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1172 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..738
; US-08-461-809-9

Query Match 3.9%; Score 55.8; DB 1; Length 1172;
Best Local Similarity 61.2%; Pred. No. 0.0088;
Matches 90; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 1288 AACATTTTCTTAAGATTTTCAAAAGGAATATGTGTAATTCAGAAATCATAACCACTG 1347
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1017 AAAGACAATTTCTATATGATGTCACATAATTTATTTAAATCAAAATTTTAGAAATG 1076
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1348 TCCTACTTGGTAAACAAACTGTTCTTAATAAAGTATTTAATGATTTTAAATGATTTTAA 1407
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1077 AATAAATTCATTCACAAAATTTAAAAAATTTAAAAAATTTAAAAAATTTAAAAAATTT 1136
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1408 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1137 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1163
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-08-461-441-9
; Sequence 9, Application US/08461441
; Patent No. 5773002
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: T CELL EPITOPES OF THE MAJOR ALLERGENS FROM
; TITLE OF INVENTION: DERMATOPHAGOIDES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII TEXT
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/461,441
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/945,288
; FILING DATE: 10 SEPTEMBER 1992
; APPLICATION NUMBER: US 580,655
; FILING DATE: 11 SEPTEMBER 1990
; APPLICATION NUMBER: US 458,642
; FILING DATE: 13 FEBRUARY 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-010CC (IMI-024)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
```



```
/ NUMBER OF SEQUENCES: 45
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Knobbe Martens Olson & Bear
/ STREET: 620 Newport Center Drive 16th Floor
/ CITY: Newport Beach.
/ STATE: California
/ COUNTRY: US
/ ZIP: 92660
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/487,826B
/ FILING DATE: 10-SEP-1993
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Israel, Ned
/ REGISTRATION NUMBER: 29,655
/ REFERENCE/DOCKET NUMBER: NIH121.001CP1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 235-8550
/ TELEFAX: (619) 235-0176
/ INFORMATION FOR SEQ ID NO: 13:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 19124 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ US-08-487-826B-13
```

```
Query Match 3.9%; Score 55.4; DB 2; Length 19124;
Best Local Similarity 48.3%; Pred.No. 0.017;
Matches 185; Conservative 0; Mismatches 196; Indels 2; Gaps 1;

QY 1052 CAATGATTAAATTTTTTTTGCATGGAGGCTTTAAATTTGTCTCTTTTCATATTTT 1111
Db 17977 CTATTTTCGATTTTTCATTTTTTTCCAGTAAATTAATTTATTTATTTTGTGATTTT 18036

QY 1112 ATTCATATTCATTTATGTTTGTACTGCTTTTGGGAGATAATTAATGTTATAAAT 1171
Db 18037 ATAATATATTTAAATGTTATTTATATATGTTTATATATGTTTATATGTTTATTTTGT 18096

QY 1172 TAGTTTGGGGGAATAATTTGTCAAAGAGGATAATTTAATTTACGTCTCTGTTATTC 1231
Db 18097 TACTCTAATCTGAATAAT--CCGAGCGAATAAATAATTAATCTCATATAAAATTA 18154

QY 1232 AGAATAAGAGAGAGACTACGCTGCATATTCAGAGTTGTACCTTAACATTTGGTGAAC 1291
Db 18155 TTTATAATACATATTTATATAGTTTCTTATTAATAAATAATTAATAATACATAATA 18214

QY 1292 ATTTTCTCAAGATTTTCAAAGGATATGTTGAATTTGAGAAATCATACCACTGCTCT 1351
Db 18215 TTCTTTGTTATTTTATAAATAAATAATTTCTTATTTTAACTTTTATTCCTTTT 18274

QY 1352 AACTTCGTAACAACTGTTCTTAAATAAGTATTTAATGATTTTAAATAAATAAATAA 1411
Db 18275 TAATTTCTTAATTTCTTTATCAACAAATAAATAAGTAATTTCTACATATCAACAAA 18334

QY 1412 AAAAAAAAAAAAAAAAAAAAAA 1434
Db 18335 AAAAAAAAAAAAAAAAAAAAAA 18357
```

```
RESULT 15
US-09-621-976-10193
; Sequence 10193, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
```

```
/ APPLICANT: Dumas Milne Edwards, J.B.
/ APPLICANT: Jobert, S.
/ APPLICANT: Giordano, J.Y.
/ TITLE OF INVENTION: ESTs and Encoded Human Proteins.
/ FILE REFERENCE: GENSET.054PR2
/ CURRENT APPLICATION NUMBER: US/09/621,976
/ NUMBER OF SEQ ID NOS: 19335
/ SOFTWARE: Patent.pm
/ SEQ ID NO 10193
/ LENGTH: 271
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ US-09-621-976-10193

Query Match 3.8%; Score 55; DB 4; Length 271;
Best Local Similarity 81.0%; Pred.No. 0.0098;
Matches 64; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1356 TGGTAAACAACTGTTCTTAAATAAAAGTATTTAAATGATTTTAAAAA 1415
Db 191 TGGGAACCAATTTTGTAAAGTAAATATTTTCACCTTTAAAAA 250

QY 1416 AAAAAAAAAAAAAAAAAA 1434
Db 251 AAAAAAAAAAAAAAAAAA 269

Search completed: July 20, 2004, 18:53:15
Job time : 131 secs
```

This Page Blank (uspto)

Result No.	Score	Query	Length	DB	ID	Description
1	1432	99-9	1434	15	US-10-195-730-12	Sequence 12, Appl
C 2	425-8	29-7	448	9	US-09-880-107-1385	Sequence 1385, Ap
3	73-2	5-1	560	17	US-10-021-323-2253	Sequence 2253, Ap
4	71-2	5-0	547	17	US-10-021-323-16890	Sequence 16890, A
C 5	70-8	4-9	520	17	US-10-021-323-7699	Sequence 7699, Ap
6	69-6	4-9	517	17	US-10-021-323-11054	Sequence 11054, A
C 7	68-4	4-8	560	17	US-10-021-323-2253	Sequence 2253, Ap
C 8	67-8	4-7	419	9	US-09-960-352-1234	Sequence 1234, A
C 9	63-6	4-4	528	17	US-10-021-323-8131	Sequence 8131, Ap
10	63-2	4-4	520	17	US-10-021-323-7699	Sequence 7699, Ap
C 11	62-4	4-4	513	17	US-10-021-323-10547	Sequence 10547, A
C 12	62-4	4-4	537	17	US-10-021-323-7228	Sequence 7228, Ap
C 13	61-8	4-3	556	17	US-10-021-323-11082	Sequence 11082, A
C 14	61-6	4-3	1284	13	US-10-434-599-25449	Sequence 25449, A

	Qy	1	CATTAAACCTTTTTTATCGGAATAGTAGATATTTTCATGTGCACCTCCATTCATGTTGA	60
	Db	1	CATTAAACTCTTTTTTATCGGAATAGTAGATATTTTCATGTGCACCTCCATTCATGTTGA	60
	Qy	61	TTTGAGCTGCACGATTATTTTGTGTAAACGACAGAGATTAATTTTATTTGAAAAGTCAGTGC	120
	Db	61	TTTGAGCTGCACGATTATTTTGTGTAAACGACAGAGATTAATTTTATTTGAAAAGTCAGTGC	120
	Qy	121	AAAAATTTATGAATAGGATATCTATAAATAACAAGATTAATACAAAAGTCGAAGCAGTGGTT	180

Query Match 4.9%; Score 70.8; DB 17; Length 520;
Best Local Similarity 49.2%; Pred. No. 0.00036;
Matches 186; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 1057 TATTAAATTTTTCATGAGGCTTTAAATTTTCGTCTTTTTCATATTTTATCA 1116
DB 450 TAAAAATATATTTTTCATCATACATTTTTCATATTTTTCATATTTTTCATCA 391
QY 1117 TATCAATTTATGTTTGTACGCTTTTTCAGGAGATAATTAATATATATTAATAGTT 1176
DB 390 AAAAAAATATATTTTTCATCATACATTTTTCATATTTTTCATATTTTTCATCA 331
QY 1177 TTGGGGGAATAATGTGCAAGAGGATAATTTAAATTTAGCTGCTTCTGTTATTCAGAT 1236
DB 330 TTTTAAATTAATTAATTAATTAATTTTATTTTATCTCTTTTTCATATTTTTCAT 271
QY 1237 AAGAGAGAGACTACCTGCATATCAAGAGTTGTACCTTAACATTTGGTGAACATTTT 1296
DB 270 AAAATACCTTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 211
QY 1297 TTCTACATTTTCAAGAGGATATGCTAAATTCAGAAATCAATACCTGCTCCTAATT 1356
DB 210 TTTAAATTTTATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 151
QY 1357 GGTAAACAACTGTTCTTAATAAGTATTTAATGATTTTAAATTTTAAATTTTAAAT 1416
DB 150 AAAAAAATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 91
QY 1417 AAAAAAATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 1434
DB 90 AAAAAAATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 73

RESULT 6

US-10-021-323-11054
; Sequence 11054, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 11054
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(517)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3828-033-Q6-N6-E7
US-10-021-323-11054

Query Match 4.9%; Score 69.6; DB 17; Length 517;
Best Local Similarity 48.9%; Pred. No. 0.00065;
Matches 186; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 1055 TGATTAATTTTTCATGAGGCTTTAAATTTTCGTCTTTTTCATATTTTAT 1114
DB 59 TTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 118
QY 1115 CATATTCATTTTTCATCTGCTTTTTCAGGAGATAATTAATATATTAATAG 1174
DB 119 TTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 178

QY 1175 TTTTGGGGGAATAATTTGTCAGAGGATAATTTAATTTACGTCTTCTGTTATTCAGA 1234
DB 179 TTTTAAAAAATAATTTTAAATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 238
QY 1235 ATAAAGAGAGAGACTACGCTGCATATTTCAAGAGTTGTACCTTAACATTTGTTGAAACAT 1294
DB 239 AAAAAAATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 298
QY 1295 TTTTCTAAGATTTTCAAAAGGATAATGTTAAATTTGAGAAATCATACCACTGCTCTAAC 1354
DB 299 AAAAAAATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 358
QY 1355 TTGTTAAACAACTGTTCTTAAATTAAGTATTTAATGATTTTAAATTTTAAATTTTAAATTTT 1414
DB 359 TTTAAATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 418
QY 1415 AAAAAAATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 1434
DB 419 TAAAAAAGGATAATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 438

RESULT 7

US-10-021-323-2253/c
; Sequence 2253, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 2253
; LENGTH: 560
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3825-011-Q1-N6-C9
US-10-021-323-2253

Query Match 4.8%; Score 68.4; DB 17; Length 560;
Best Local Similarity 48.9%; Pred. No. 0.0012;
Matches 183; Conservative 0; Mismatches 191; Indels 0; Gaps 0;

QY 1061 AATTTTTCATGAGGCTTTAAATTTTCGTCTTTTTCATATTTTATTCATATTT 1120
DB 374 ATTTTTCATGAGGCTTTAAATTTTCGTCTTTTTCATATTTTATTCATATTT 315
QY 1121 CAATTTATGTTTGTAACTTTTTCAGGAGATAATTTATATGTTAATTTAGTTTTCG 1180
DB 314 TTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 255
QY 1181 GGGGAATAATTTGTCAGAGGATAATTTAAATTTACGTCTTCTGTTATTCAGAAATAAG 1240
DB 254 TAATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 195
QY 1241 AGAAGAGACTACGCTGCATATTTCAAGAGTTGTACCTTAACATTTGTTGAAACATTTTCT 1300
DB 194 ATAAATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 135
QY 1301 AAGATTTTCAAAAGGATAATGTTAAATTTGAGAAATCATACCACTGCTCTAACTTGGTA 1360
DB 134 AAAAAAATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCATATTTTTCAT 75
QY 1361 AACAACTGTTCTTAATTAAGTATTTAATGATTTTAAATTTTAAATTTTAAATTTTAAAT 1420

```

Db      74  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 15
Qy      1421 AAAAAAAAAAAAAAAAAAAAA 1434
         ||||| |||||
Db      14  AAAAAACAAAAAAA 1

```

RESULT 8

```

RESULT 8
US-09-960-352-11234/c
; Sequence 11234, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11234
; LENGTH: 419
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-LIB3058-052-Q1-K1-D8
US-09-960-352-11234

```

	Query Match	4.7%;	Score 67.8;	DB 9;	Length 419;
	Best Local Similarity	51.0%;	Pred.No. 0.0013;		
	Matches 187;	Conservative 0;	Mismatches 177;	Indels 3;	Gaps 1;
Qy	1068	TTTTGCATGAGGGCTTTAAAAATTGGCTCTTTTTCATATTTTATTCATATTCAAATTA	1127		
Db	415	TTTTTATAAATATATATTTAAAAAATAAATAATTTATATAAAAAATTTTTTATAAAA	356		
Qy	1128	TGCTTTTGTAACGTCTTTTAGGGAGATAATATATATGTTATAAAATTTAGTTTGGGGGAAT	1187		
Db	355	ATAATTTTTTAAAAAATTTTTTTTAAAAATTTTTTATTTATTTATTTTAAATATAATT	296		
Qy	1188	AATTGTCCAAGAGGATATATTAATTTTACGTGCTCTGTATTTCAGATAAAGACAGAG	1247		

RESULT 9

```

RESULT 9
US-10-021-323-8131/c
; Sequence 8131, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants

```

```

; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US
; CURRENT FILING DATE: 2001-12
; PRIOR APPLICATION NUMBER: US
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 8131
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
;
; OTHER INFORMATION: Clone ID:
US-10-021-323-8131

```

	Query Match	4.4%;	Score 63.6;	DB 17;	Length 528;
	Best Local Similarity	48.9%;	Pred. No. 0.01;		
	Matches 171;	Conservative 0;	Mismatches 179;	Indels 0;	Gaps 0;
Qy	1082	CTTTAAAAATTGCGCTCTTTTTTCATATTTTATTCATATTCAAATTATGTGGTTTGTAACATGC	1141		
Dd	350	CCTAAAAAATTTTTTTTTTTTTTTTTTTTTTTTAAAACCCCTTTTTTTTTTTTTTTTTTTTTTTT	291		
Qy	1142	TTTTTAGGAGATAAATTATATGTATATAAATTAGTTTGGGGGAATAAATTGTGCAAGAAG	1201		
Dd	290	TTTAAACCCTTTTTTTTTTTTTTT	231		
Qy	1202	GATAAATTAATTTACGTGCTTCTGTATTTCAGATAATAAGAGAGAAGACTAGGCTGCATAT	1261		
Dd	230	TTTAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTAAAAAAAACCCCCCTTTTTTTTTT	171		
Qy	1262	TCAAGAGTTGTACCTTAACAATGGTGAACATTTTTCCTAAGATTTTTCAAAAAGGAATATG	1321		
Dd	170	TTTTTTTTTTTTTTTTTAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	111		
Qy	1322	TGTAAAATTGAGAATCATAAACCACTGCTCAAATGGTGAACCAACTGTTCTTAAATAAA	1381		
Dd	110	AAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	51		
Qy	1382	GTATTTTAATCATTTTAAAAAANAANAANAANAANAANAANAANAANAANAANA	1431		
Dd	50	AAAAAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANAANA	1		

RESULT 10

```

RESULT 10
US-10-021-323-7699
; Sequence 7699, Application US/10021323
; Publication No. US20040123340A1
; GENERAL INFORMATION:
; APPLICANT: Deikman, Jill
; APPLICANT: Feng, Paul C.C.
; APPLICANT: Fincher, Karen L.
; APPLICANT: Ziegler, Todd E.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(52274)B
; CURRENT APPLICATION NUMBER: US/10/021,323
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: US 60/255, 619
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 17880
; SEQ ID NO 7699
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: LTB3828-019-Q1-N6-D6
US-10-021-323-7699

```

Query Match 4.4%; Score 63.2; DB 17; Length 520;
Best Local Similarity 48.1%; Pred. No. 0.012;
Matches 179; Conservative 0; Mismatches 193; Indels 0; Gaps 0;
Ov 1063 TTTTTTTTTGCAGAGGGCTTAAAAATTTGTCTCTTTTCATATTTATCATATTTCA 1122

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 17:01:24 ; Search time 3467 Seconds
(without alignments)

12351.418 Million cell updates/sec

Title: US-10-799-747-12

Perfect score: 1434

Sequence: 1 cattaaactcttttatcg.....aaaaaaaaaaaaaaaaaaaaa 1434

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	871.2	60.8	1201	13 BX441923	BX441923 BX441923
2	744.6	51.9	814	14 CB853053	CB853053 UI-CF-FNO
3	643.6	44.9	765	12 B1914473	B1914473 603182264
4	606	42.3	785	12 BG484396	BG484396 602505037

C	5	576.2	40.2	588	10	BF431622	BF431622 7016e08.x
C	6	551.4	38.5	568	9	AL035942	AL035942 DRFP564B
	7	549.8	38.3	575	12	BM722991	BM722991 UI-E-E01-
	8	543.6	37.9	557	14	CB132625	CB132625 K-EST0183
C	9	537.2	37.5	556	10	AW500190	AW500190 UI-HF-BNO
	10	534	37.2	551	12	B1259796	B1259796 602971440
	11	513.8	35.8	621	9	AL035941	AL035941 DRFP564B
	12	468.2	32.6	503	10	AW889139	AW889139 QV4-NT002
C	13	444	31.0	444	9	AA136080	AA136080 zK90D03.s
	14	427	29.8	438	10	AW304923	AW304923 xv97b07.x
C	15	425.8	29.7	448	9	AA491000	AA491000 aa52g11.s
C	16	415.2	29.0	1037	13	EX413464	EX413464 BX413464
C	17	396.4	27.6	400	9	AW002842	AW002842 wt56h11.x
C	18	395	27.5	395	9	AA92811	AA92811 wt56h11.x
C	19	372.6	26.0	688	12	BM728856	BM728856 UI-E-E01-
C	20	370.6	25.8	394	9	AA747222	AA747222 nx76a10.s
C	21	354.6	24.7	419	10	AW189910	AW189910 x110h04.x
C	22	349.4	24.4	493	9	AW026264	AW026264 wt10h02.x
	23	323.8	22.6	941	13	EX452125	EX452125 BX452125
	24	295.4	20.6	957	12	BG403642	BG403642 602419361
C	25	278.4	19.4	280	9	AA679646	AA679646 ah10c06.s
C	26	273	19.0	338	12	BI025209	BI025209 CM4-NT028
C	27	233	16.2	253	9	AA953459	AA953459 on69g09.s
C	28	168.8	11.8	240	10	BF910533	BF910533 CM4-UT004
	29	88.6	6.2	1084	13	EX361825	EX361825 BX361825
C	30	84.6	5.9	1126	13	EX446388	EX446388 EX446388
C	31	84.4	5.9	1101	29	CNS00EQL	AL069526 Drosophil
C	32	83.6	5.8	1133	13	EX444099	EX444099 BX444099
	33	82.4	5.7	634	13	BX506904	BX506904 DRFP2779F
	34	81.6	5.7	1056	13	EX415058	EX415058 BX415058
	35	81.4	5.7	1101	29	CNS00EVL	AL069706 Drosophil
	36	81	5.6	1200	13	EX437758	EX437758 BX437758
	37	80.4	5.6	625	29	CNS036A2	AL229763 Tetraodon
	38	80.4	5.6	1200	13	EX436510	EX436510 BX436510
C	39	79.4	5.5	1043	29	CNS0145P	AL103735 Drosophil
	40	78.6	5.5	681	13	EX378802	EX378802 BX378802
C	41	77.8	5.4	1101	29	CNS00EVL	AL069706 Drosophil
	42	77.8	5.4	1201	13	EX402521	EX402521 BX402521
	43	77.4	5.4	450	29	CNS020OX	AL207042 Tetraodon
	44	76.6	5.3	999	13	BX380865	BX380865 BX380865
C	45	76	5.3	681	13	BX378802	BX378802 BX378802

ALIGNMENTS

RESULT 1
BX441923

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BX441923 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF023Y002 5-PRIME, mRNA sequence.
BX441923.1 GI:30771989

EST.

Homo sapiens (human)

Homo sapiens

1 (bases 1 to 1201)

Li M.B., Gruber C., Jessee J. and Polayes D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 2532.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DF023BH01QPI&cluster=2532.r. Contact :

Feng Liang Email: fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0DF023BH01QPI.

FEATURES

source

```
1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF023Y002"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
```

ORIGIN

```
Query Match      60.8%; Score 871.2; DB 13; Length 1201;
Best Local Similarity 93.0%; Pred. No. 2.3e-104;
Matches 937; Conservative 21; Mismatches 37; Indels 13; Gaps 4;

QY 415 CTTCTAGGTAGGGAGCTTATTCAGACATAGTAGATTTCTCTAATGCTGTSTCAAT 474
Db |||||
QY 54 CTTCTAGGTAGGGAGCTTATTCAGACATAGTAGATTTCTCTAATGCTGTCTCAAT 113
Db |||||
QY 475 TGCTGGCCTTTGGCTACCTGCTACTTCCSCATTATGGCAGCCCATTCAGTCTTGAGTTTC 534
Db |||||
QY 114 TGCTGGCCTTTGGCTACCTGCTACTTCCSCATTATGGCAGCCCATTCAGTCTTGAGTTTC 173
Db |||||
QY 535 TTCTCTGACACCTTATGCTCTGAATCATGAGGAGCTGATTCATTTGGTGAATTTGGG 594
Db |||||
QY 174 TTCTCTGACACCTTATGCTCTGAATCATGAGGAGCTGATTCATTTGGTGAATTTGGG 233
Db |||||
QY 595 TAGAAAGCAGTATGTTTGTGACATTAAGTGTAGTATAGATAGTTTGTAGCTTTAA 654
Db |||||
QY 234 TAGAAAGCAGTATGTTTGTGACATTAAGTGTAGTATAGATAGTTTGTAGCTTTAA 293
Db |||||
QY 655 GTGTATGTTTTATACCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCACCTCTCC 714
Db |||||
QY 294 GTGTATGTTTTATACCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCACCTCTCC 353
Db |||||
QY 715 CCCAGCCTATCTCAAACTGGTGGATATATGAGAGATCTTTGAAAGAGTAAATAAAC 774
Db |||||
QY 354 CCCAGCCTATCTCAAACTGGTGGATATATGAGAGATCTTTGAAAGAGTAAATAAAC 413
Db |||||
QY 775 TTCACTGCTCCACTCCAGGTAATCCGCCACTCCCACTGACCTAGTAGATTTGTAAT 834
Db |||||
QY 414 TTCACTGCTCCACTCCAGGTAATCCGCCACTCCCACTGACCTAGTAGATTTGTAAT 473
Db |||||
QY 835 TAATACTTACTCTTATTTCTGAATCAGTTGTGAATGTTGCTTATGTTTCAGAGTTT 894
Db |||||
QY 474 TAATACTTACTCTTATTTCTGAATCAGTTGTGAATGTTGCTTATGTTTCAGAGTTT 533
Db |||||
QY 895 AAGAACCTGCTGAATTCATTTTAAATCTGCTATCTGAGAGCAATGAATGAATTC 954
Db |||||
QY 534 AAGAACCTGCTGAATTCATTTTAAATCTGCTATCTGAGAGCAATGAATGAATTC 593
Db |||||
QY 955 TTAACAAGAGAGACTCATCTGTAGCTGTTTGTGACTCTATGAGCCCAATAGGTTCTG 1014
Db |||||
QY 594 TTAACAAGAGAGACTCATCTGTAGCTGTTTGTGACTCTATGAGCCCAATAGGTTCTG 652
Db |||||
QY 1015 TGCTTAGCATTAACAAATAAGTTTATAGTTAAAGCCCAATGATTTAA-TTTTTTTTTTC 1073
Db |||||
QY 653 TGCTTAGCATTAACAAATAAGTTTATAGTTAAAGCCCAATGATTTAA-TTTTTTTTTTC 712
Db |||||
QY 1074 ATGAGGGCTTTAAATTTGCTGCTTTTTCATATTTTATTCATTTTCAATTTATGTTT 1133
Db |||||
QY 713 ATGAGGGCTTTAAATTTGCTGCTTTTTCATATTTTATTCATTTTCAATTTATGTTT 772
Db |||||
QY 1134 GTAACCTGTTTTTAGGGAGATAATATATGTTTAAATAGTTTGTGGGGGGAATAATGT 1193
Db |||||
QY 773 GTAACCTGTTTTTAGGGAGATAATATATGTTTAAATAGTTTGTGGGGGGAATAATGT 832
Db |||||
```

```
QY 1194 GCBAAGAGGATAATTTAATTTACGTGCTTCTGTTATTTCAGATAAAGAGAGACTACG 1253
Db |||||
QY 833 GCBAAGAGGATAATTTAATTTACGTGCTTCTGTTATTTCAGATAAAGAGAGACTACG 892
Db |||||
QY 1254 CTGCATATTTCAAGAGTTGCTACCTTAACATTTGCTGTAACATTTTCTTAAGATTTTCAAA 1313
Db |||||
QY 893 CTGCATATTTCAAGAGTTGCTACCTTAACATTTGCTGTAACATTTTCTTAAGATTTTCAAA 951
Db |||||
QY 1314 GGAATATGTTAAATTTGAGAATCATACCACTGCTCTAATCTTGTGTAACAACTGTTCT 1373
Db |||||
QY 952 GAAATGCTGTTTAAATTTGAGAATCATACCACTGCTCTAATCTTGTGTAACAACTGTTCT 1001
Db |||||
QY 1374 TAAATAAGTATTTAATGATTTTAAATTTGAGAATCATACCACTGCTCTAATCTTGTGTAACAACTGTTCT 1421
Db |||||
QY 1002 CTAATAAGTATTTAATGATTTTAAATTTGAGAATCATACCACTGCTCTAATCTTGTGTAACAACTGTTCT 1049
Db |||||
```

RESULT 2

CB853053/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

```
CB853053      814 bp      mRNA      linear      EST 22-APR-2003
UI-CF-FNO-afg-e-20-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone
UI-CF-FNO-afg-e-20-0-UI 3', mRNA sequence.
CB853053
CB853053.1 GI:30048171
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 814)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com)
The following repetitive elements were found in this cDNA
sequence: 1-42, >AT-rich#low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.
```

FEATURES

source

```
1. 814
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-FNO-afg-e-20-0-UI"
/tissue_type="Human Lung Epithelial cells"
/lab_host="DH10B (Life Technologies)" (T1 phage resistant)"
/clone_lib="UI-CF-FNO"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-FNO is a subtracted cDNA library derived from two
normalized Human lung epithelial cell libraries (EN1 and
DUI) The library was subtracted according to according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. For additional information, contact:
bento-soares@uiowa.edu
TAG_SEQ=None found"
```

ORIGIN

[illegible]

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 765)
NIH-MGC <http://imgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11621 row: p column: 16
High quality sequence stop: 762.

FEATURES source.

1. 765
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5246463"
/lab_host="DH10B"
/clone_lib="NIH_MGC_121"
/note="Organ: brain; Vec
Site 2: EcoRV (destroyed
fetal brains, female age
and male age 26 weeks.
directionally cloned (Ec
cloning). Average inser
0.7-3.5 kb. Library is
full-length clones and
(Invitrogen). Research
this is a NIH MGC Library

ORIGIN

Query March	44.9%	Score 643.6	DB 12	Length 765
Best Local Similarity	97.8%	Pred. No. 1.1e-74		
Matches 702	Conservative 3	Mismatches 7	Indels 6	Gaps 5
QY	1	CATTAAACTCTTTTTATCGGGAATAGTATGATATATTTCAATGTCACATCCATTTCATGTTGA	60	
Db	49	CAITPAAACTCTTTTTATCGGGAATAGTATGATATATTTCAATGTCACATCCATTTCATGTTGA	107	
QY	61	TTTCGAGCTGACAGCTTATTTTGTGTAAGCAGAGATTTAAATTTTATATGTGAAAGTCAGTGC	120	
Db	108	TTTCGAGCTGACAGCTTATTTTGTGTAAGCAGAGATTTAAATTTTATATGTGAAAGTCAGTGC	167	
QY	121	AAAATTTATGATAGATATACTAATTAATACAAAGTAATACAAAAGTCAAAAGCAGTGTT	180	
Db	168	AAAATTTATGATAGATATACTAATTAATACAAAAGTAATACAAAAGTCAAAAGCAGTGTT	227	
QY	181	CTAAATAAAATCTCGGTTCCCTTAAAAAATTTATTTTAAATTTATC-TTGAATAGTTTTC	239	
Db	228	CTAAATAAAATCTCGGTTCCCTTAAAAAATTTATTTTAAATTTATCTTTGAAATAGTTTTC	287	
QY	240	TTAGATTAATCTCAGGATATCAGAAAAGTCAAATTAAGTGTGAGTAAAGTTAGTATCATTTAA	299	
Db	288	TTAGATTAATCTCAGGATATCAGAAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA	347	
QY	300	ACAAATGTCTATTAATATGCAMGA-GTGGTAATATACAGAATTTATCAGGCATTACCAAG	358	
Db	348	ACAAATGTCTATTAATATGCAMGACGTGGTAATATACAGAATTTATCAGGCATTACCAAG	407	
QY	359	TCTAGGCACATATAGGAAATGCAGCACTCAGAAATGGTTTCAATCTAGTAGTTGATGCTTG	418	
Db	408	TCTAGGCACATATAGGAAATGCAGCACTCAGAAATGGTTTCAATCTAGTAGTTGATGCTTG	467	
QY	419	TAAAGTAGGGAGCTTATTCAGACATAGTAGATAGTTTCTCTAATGCTGTGTCAATTGCT	478	
Db	468	TAAAGTAGGGAGCTTATTCAGACATAGTAGATAGTTTCTCTAATGCTGTGTCAATTGCT	527	
QY	479	GGCCTTTTGGCTACCTGTACTTCCSACATTATGGACCCCAATTCAGTCTTGAGTTTTCTTCT	538	

Db 528 GGCCTTTGGCTACCTGTACTTCCCATTAATGACGAGCCATTGAGTTTCTCT 587
 QY 539 CTGGACACCTTATGCTCTGAATCATGAGCGAGGCTGATTCATTAATTTGGGTAGA 598
 Db 588 CTGGACACCTTATGCTCTGAATCATGAGCGAGGCTGATTCATTAATTTGGGTAGA 646
 QY 599 AAGCAGTAGTGTCTGACATTAAGATGTAGTTATAGATAGGTTTAGCCCTTTAAGTGT 658
 Db 647 AAGCAGTAGTGTCTGACATTAAGATGTAGTTATAGATAGGTTTAGCCCTTTAAGTGT 706
 QY 659 ATGTTTTTATCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCCACCTCTCTCCCC 716
 Db 707 ATG--TCTATACTTTAAATAAGAAATATAACCTTTTAAAGCTATTCCACCTCTCTCCCC 762

RESULT 4
 BG484396 785 bp mRNA linear EST 21-MAR-2001
 LOCUS 6025037F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4618473 5',
 DEFINITION mRNA sequence.

ACCESSION BG484396
 VERSION BG484396.1 GI:13416675

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: CLONTECH Laboratories, Inc.

cDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LICM1376 row: j column: 10

High quality sequence start: 18

High quality sequence stop: 751.

Location/Qualifiers

FEATURES

source

1..785

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="IMAGE:4618473"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NTH_MGC_77"

/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site:1:

Sfil (ggccgctcggcc); Site:2: Sfil (ggccattatggcc); 5' and

3' adaptors were used in cloning as follows: 5' adaptor

sequence: 5'-CAGCGCATTAATGCCC-3' and 3' adaptor sequence:

5'-ATTAGAGCGCGGCGGCGCATG-dt(30)BN-3' (where B = A,

C, or G and N = A, C, G, or T). Average insert size 1.9

kb (range 0.5-4.0 kb). 12/15 colonies contained inserts

by PCR. This library was enriched for full-length clones

and was constructed by Clontech laboratories (Palo Alto,

CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 42.3%; Score 606; DB 12; Length 785;

Best Local Similarity 95.4%; Pred. No. 7.9e-70;

Matches 682; Conservative 5; Mismatches 21; Indels 7; Gaps 6;

QY 249 TCTCAGGATATGAGAAAGTCAATTAAGTGTAGTGAATAGTATATCAATTAACAAATGT 308

Db 63 TCTAGAGATATGACACCGTCGAGTAAGTGTGAGTAAAGTTAGTATCAATTAACAAATGT 122

QY 309 CTATTAATGCMGA-GTGGTGAATATACAGAAATTTATCAGGCATTACCAAGCTAGGCAC 367

Db 123 CTATTAATGCMGACGCTGGTGAATATACAGAAATTTATCAGGCATTACCAAGTCTAGGCAC 182
 QY 368 ATATAGGAAATGCGAGCACTCAGAATGGTTTCAATGCTAGTGTGCTGTGTAAGGTAGG 427
 Db 183 ATATAGGAAATGCGAGCACTCAGAATGGTTTCAATGCTAGTGTGCTGTGTAAGGTAGG 242
 QY 428 GGAGCTTATTGACATATAGTATAGTATTTCTCTAATGCTGTSTCAATTTGCTGGCCTTTGG 487
 Db 243 GGAGCTTATTGACATATAGTATAGTATTTCTCTAATGCTGTCTCAATTTGCTGGCCTTTGG 302
 QY 488 CTACTGTACTTCCSCATTATGGCAGCCCAATTCAGTCTTGTAGTTTCTCTCTGACACCC 547
 Db 303 CTACTGTACTTCCSCATTATGGCAGCCCAATTCAGTCTTGTAGTTTCTCTCTGACACCC 362
 QY 548 TTATGCTCTGAAATCATGAGCGAGGCTGATTCATTAATGCTGTATTTGGGTAGAAAGCAGTAT 607
 Db 363 TTATGCTCTGAAATCATGAGCGAGGCTGATTCATTAATGCTGTATTTGGGTAGAAAGCAGTAT 422
 QY 608 GTTTTGTCTGACATTAAGATGTAGTATAGATAGTATTTAGCTTTAAGTGTATGTTTTTA 667
 Db 423 GTTTTGTCTGACATTAAGATGTAGTATAGATAGTATTTAGCTTTAAGTGTATGTTTTTA 482
 QY 668 TACTTTAAATAAGAAATATATAACCTTTTAAAGCTATTCACCTCTCTCCCGCAGCCTATCTC 727
 Db 483 TACTTTAAATAAGAAATATATAACCTTTTAAAGCTATTCACCTCTCTCCCGCAGCCTATCTC 542
 QY 728 AAACCTGGTGAATATATGGAGAGATCTTGAAGAAGTAAATAAAACCTTCACTGCTCCAC 787
 Db 543 AAACCTGGTGAATATATGGAGAGATCTTGAAGAAGTAAATAAAACCTTCACTGCTCCAC 602
 QY 788 TCCAGGTGAATCCCGCCACTCCCACTGACCTAGTAGTAA-TTTGTAATTTAATTAATTAAT 846
 Db 603 TCCAGGTGAATCCCGCCA-TCCCACTGACCTAGTAGTAA-TTTGTAATTTAATTAATTAAT 661
 QY 847 TCTATTTCTGAAATCAGTTGTGAAT-CTTGCCCTTATGTTTTCAGAGTGTTAAGAACCCTCMG 905
 Db 662 TCTA-TTCTGAAATCAGTTGTGAATCTTGTTGCTTATGTTTCAGAGTGTTAAGAACCCTCAG 720
 QY 906 TGAATTCATTTTAAATCTGCTA--TTCTGAGAAGCATTGAATGAATTTCTTAA 958
 Db 721 TGAATTCATTTTAAATCTGCTAATTTCTGAGGAGGCAATGACTGAATTTCTTAA 775

RESULT 5

BF431622/c

LOCUS 7016e08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3574359 3',

DEFINITION mRNA sequence.

ACCESSION BF431622

VERSION BF431622.1 GI:11443736

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 588)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Unpublished (1997)

JOURNAL Tumor Gene Index

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Sequencing by: Greg Lemmon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

infoimage.llnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 469.

FEATURES	source	Location/Qualifiers
		1. .588
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone="IMAGE:3574359"
		/lab_host="DH10B"
		/lab_lib="NCI CGAP Kid11"
		/note="Organ: kidney; Vector: pTV73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Plasmid DNA from the normalized library NCI CGAP Kid3 was prepared, and ss circles were made in vitro. Following RAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clone)Ds 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Patricia Bonaldo."
ORIGIN		
Query Match	40.2%;	Score 576.2; DB 10; Length 588;
Best Local Similarity	99.5%;	Pred. No. 7.1e-66;
Matches 586;	Conservative 2;	Mismatches 0; Indels 1; Gaps 1;
Qy	812	CTGACCTAGTAGAATTTGTAAATTAACCTACCTCTCTATTTCTGAAATCAGTTGTGAAC 871
Db	588	CTGACCTAGTAGAATTTGTAAATTAACCTACCTCTCTATTTCTGAAATCAGTTGTGAAC 529
Qy	872	TGTTGCCCTATGTCAGAGCTTTAAGAACCTCMGTGAATTCATTTTTTAAATCTGCTAT 931
Db	528	TGTTGCCCTATGTTACAGAGTTTAAAGAACCTCAGTGAATTCATTTTTTAAATCTGCTAT 469
Qy	932	TCTGAGAAGCATTTGAATGAATTCCTTAACAAGAGACTCATCTGTAGCTGTTTGTCTGACTC 991
Db	468	TCTGAGAAGCATTTGAATGAATTCCTTAACAAGAGACTCATCTGTAGCTGTTTGTCTGACTC 409
Qy	992	CTATGAGCCCATAAAGGGTTCTGTGCTTAGCATTAACAAATAAGGTTTATAGTAAAGC 1051
Db	408	CTATGAG-CCCATAAAGGGTTCTGTGCTTAGCATTAACAAATAAGGTTTATAGTAAAGC 350
Qy	1052	CAATGTTATTAATTTTTTTTGCATGGAGGGCTTTAAATTTTGCTCTCTTTTCATATTTT 1111
Db	349	CAATGTTATTAATTTTTTTTGCATGGAGGGCTTTAAATTTTGCTCTCTTTTCATATTTT 290
Qy	1112	ATTCAATATTCAAATTTATGGTTTGTAACTGCTTTTACGGAGATAATTTATGTTATATAAT 1171
Db	289	ATTCAATATTCAAATTTATGGTTTGTAACTGCTTTTACGGAGATAATTTATGTTATATAAT 230
Qy	1172	TAGTTTTGGGGGAATTAATTTGTCAAAAGAGGATAAATTTAATTTACGTGCTTCTGTTATTC 1231
Db	229	TAGTTTTGGGGGAATTAATTTGTCAAAAGAGGATAAATTTAATTTACGTGCTTCTGTTATTC 170
Qy	1232	AGATAAAGAGAGAGACTACGCTGCGATATTCAGAGTTGTACCTTAACATTGGTGAAC 1291
Db	169	AGAATAAAGAGAGAGACTACGCTGCGATATTCAGAGTTGTACCTTAACATTGGTGAAC 110
Qy	1292	ATTTTTTCTAAGATTTTCAAAGGAATATGTGAATTTGAGAAATCATACCACTGTCCT 1351
Db	109	ATTTTTTCTAAGATTTTCAAAGGAATATGTGAATTTGAGAAATCATACCACTGTCCT 50
Qy	1352	AACCTGGTAAACAACTGTTCTTAAATAAGTATTTAATGATTTTAAAA 1400
Db	49	AACCTGGTAAACAACTGTTCTTAAATAAGTATTTAATGATTTTAAAA 1
RESULT 6		
AL035942/c		
LOCUS	AL035942	568 bp mRNA linear EST 04-SEP-2003
DEFINITION	DFXP2564B1622 s1 564 (synonym: hfbz2) Homo sapiens cDNA clone	
	DFXP2564B1622 3', mRNA sequence.	
ACCESSION	AL035942	
VERSION	AL035942.1	GI:5405576
KEYWORDS	EST.	

```

QY 1384 ATTTAATGATTTTAAAAAAGAAAAA 1412
Db 29 ATTTAATGATTTTAAAAAAGAAAAA 1

RESULT 7
BM722991
LOCUS BM722991
DEFINITION UI-E-Eol-aid-a-14-0-UI.r1 UI-E-Eol Homo sapiens cDNA clone
ACCESSION BM722991
VERSION UI-E-Eol-aid-a-14-0-UI 5', mRNA sequence.
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 575)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 148-169, >AT-rich#Low_complexity (matched complement)
Seq primer: M13 Reverse.
Location/Qualifiers
1..575
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-Eol-aid-a-14-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-Eol"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-Eol is a normalized cDNA library containing the
following tissue(s): fetal eye. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CGCGTATACC. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI)."
```

Query Match 38.3%; Score 549.8; DB 12; Length 575;
Best Local Similarity 99.1%; Pred. No. 1.9e-62;
Matches 570; Conservative 3; Mismatches 0; Indels 2; Gaps 2;

QY 56 GTTGAATTTGGAGCTGACAGTATTTTGTGTAGCAGAGATTTAATTTATATTGAAGTC 115

```

Db 1 GTTGAATTTGGAGCTGACAGTATTTTGTGTAGCAGAGATTTAATTTATATTGAAGTC 60
QY 116 AGTGCAAAATTTATGAATAGGATATATACTAATAATAACAAGTAATAACAAAAGTCAAGCA 175
Db 61 AGTGCAAAATTTATGAATAGGATATATACTAATAATAACAAGTAATAACAAAAGTCAAGCA 120
QY 176 GTGTTCTAAATAAATAATTTCTGGGTTCCCTTAAATAATTTTAAATTTATTC-TTGAATAAG 234
Db 121 GTGTTCTAAATAAATAATTTCTGGGTTCCCTTAAATAATTTTAAATTTATTC-TTGAATAAG 180
QY 235 TTTTCTTAGATTAAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTGTAGTATC 294
Db 181 TTTTCTTAGATTAAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTGTAGTATC 240
QY 295 ATTAACAAATTTGTCTATTAAATGCMGA-GTGTGTAATATACAGAAATTTATCAGGCATT 353
Db 241 ATTAACAAATTTGTCTATTAAATGCMGACGCTGGTAAATATACAGAAATTTATCAGGCATT 300
QY 354 CCAAGTCTAGGCACATATAGGAATGCGACCTCAGAAATGCTTCAATGTAGTATTGAT 413
Db 301 CCAAGTCTAGGCACATATAGGAATGCGACCTCAGAAATGCTTCAATGTAGTATTGAT 360
QY 414 GCTTGTAAAGTAGGGAGCTTATTTCAGACATAGTAGATAGTTCTCTAATGCTGTSTCAA 473
Db 361 GCTTGTAAAGTAGGGAGCTTATTTCAGACATAGTAGATAGTTCTCTAATGCTGTCTCAA 420
QY 474 TTGCTGGCCTTTGGCTACCTGCTACCTCCGCAATTTAGCGAGCCCATTCAGTCTGAGTTT 533
Db 421 TTGCTGGCCTTTGGCTACCTGCTACCTCCGCAATTTAGCGAGCCCATTCAGTCTGAGTTT 480
QY 534 CTCTCTCGGACACCTTATGCTCTGAATCATGAGGAGGCTGATTCATTTGGTGGTTGG 593
Db 481 CTCTCTCGGACACCTTATGCTCTGAATCATGAGGAGGCTGATTCATTTGGTGGTTGG 540
QY 594 GTAGAAACAGTATGTTTGTGCTGACATTAAGATGT 628
Db 541 GTAGAAACAGTATGTTTGTGCTGACATTAAGATGT 575

RESULT 8
CB132625
LOCUS CB132625
DEFINITION K-EST0183064 L14ChoiCK0 Homo sapiens cDNA clone L14ChoiCK0-2-D11
5', mRNA sequence.
ACCESSION CB132625
VERSION CB132625.1 GI:28097898
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 557)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 2 row: D column: 11
High quality sequence stop: 557.
Location/Qualifiers
1..557
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L14ChoiCK0-2-D11"
```


/sex="M"
 /cell_line="Choi-Ck"
 /lab_host="Top10F"
 /clone_lib="L14ChoiCK0"
 /notes="Organ: Liver; Vector: pT73-Pac; Site 1: EcoRI;
 Site 2: NotI; The library was contributed by the Soares
 laboratory and it was constructed as described by Bonaldo,
 M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
 6(9): 791-806. RNA was prepared from harvested cell
 culture."

ORIGIN

Query Match 37.9%; Score 543.6; DB 14; Length 557;
 Best Local Similarity 99.3%; Pred. No. 1.3e-61;
 Matches 554; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY 674 AATAAGAAATTAACCTTTTAACTTATTCACCTCTCCCTCCAGCCTATCTCAACTG 733
 Db 1 AATAAGAAATTAACCTTTTAACTTATTCACCTCTCCCTCCAGCCTATCTCAACTG 60
 QY 734 GTGGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCACCTGCTCCACTCCAGG 793
 Db 61 GTGGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCACCTGCTCCACTCCAGG 120
 QY 794 TGAATCGCCCACTCCCACTGACCTAGTAGAATTTGTAATTAATTAATTAATTAATTTT 853
 Db 121 TGAATCGCCCACTCCCACTGACCTAGTAGAATTTGTAATTAATTAATTAATTTT 180
 QY 854 CTGAATCACTGTGACCTGTGCTTATCTTCAAGAGTTTAAAGAACTCMTGTAATTC 913
 Db 181 CTGAATCACTGTGACCTGTGCTTATCTTCAAGAGTTTAAAGAACTCMTGTAATTC 240
 QY 914 TTTTAAAAATCTGCTATCTGAGAAGCAITGAATGAATTTTAAACAAGAGCTCATCT 973
 Db 241 TTTTAAAAATCTGCTATCTGAGAAGCAITGAATGAATTTTAAACAAGAGCTCATCT 300
 QY 974 GTAGCTGTTTGTGACCTCTATGAGCCCAATAGAGGTTCTGTGCTTAGCAATTAACAAAT 1033
 Db 301 GTAGCTGTTTGTGACCTCTATGAG-CCCAATAGAGGTTCTGTGCTTAGCAATTAACAAAT 359
 QY 1034 AAGTTTATAGTAAAGCCCAATGATTAATTTTGTGATGAGGCTTTAAATTTG 1093
 Db 360 AAGTTTATAGTAAAGCCCAATGATTAATTTTGTGATGAGGCTTTAAATTTG 419
 QY 1094 TGCTCTTTTCAATTTTATTCATATTCATATTCATTTATGTTTGTGTAAGGAGA 1153
 Db 420 TGCTCTTTTCAATTTTATTCATATTCATTTATGTTTGTGTAAGGAGA 479
 QY 1154 TAATTAATGTTATAAATTTAGTTTGGGGGAATAATTTGCAAGAAGATTAATTTAAT 1213
 Db 480 TAATTAATGTTATAAATTTAGTTTGGGGGAATAATTTGCAAGAAGATTAATTTAAT 539
 QY 1214 TACGTGCTTCTGTTATTC 1231
 Db 540 TACGTGCTTCTGTTATTC 557

RESULT 9
 AW500190/c
 LOCUS 556 bp mRNA linear EST 01-MAR-2000
 DEFINITION UI-HF-BNO-akc-g-07-0-UI.r1 NIH_MGC_50 Homo sapiens cDNA clone
 IMAGE:3076789 5', mRNA sequence.
 ACCESSION AW500190
 VERSION AW500190.1 GI:7112570
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 556)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: M.B. Soares Lab

CDNA Library Arrayed by: M.B. Soares Lab

DNA Sequencing by: M.B. Soares Lab

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

The following repetitive elements were found in this cDNA sequence:

8-48, >AR_rich#Low_complexity
 Seq primer: M13 Forward.

FEATURES

source

1..556
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:3076789"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LTI)"
 /clone_lib="NIH_MGC_50"
 /notes="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI;
 Constructed from size fractionated cytoplasmic mRNA
 (3.5-4.4kb). Directionally cloned. Cells provided by
 Louis M. Staudt, Ph.D. Library preparation by Maria de
 Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

ORIGIN

Query Match 37.5%; Score 537.2; DB 10; Length 556;
 Best Local Similarity 99.5%; Pred. No. 8.5e-61;
 Matches 547; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 864 TTGTGAACCTGTCCTTATGTTTCTGAGAGTTTAAAGAACTCMTGGAATTCATTTTAAAA 923
 Db 556 TTGTGAACCTGTCCTTATGTTTCTGAGAGTTTAAAGAACTCAGTGAATTCATTTTAAAA 497
 QY 924 TCTGCTATTCTGAGAGCAITGAATGAATTTCTTAAACAGAGAGCTCATCTGTAGCTGTTT 983
 Db 496 TCTGCTATTCTGAGAGCAITGAATGAATTTCTTAAACAGAGAGCTCATCTGTAGCTGTTT 437
 QY 984 GCTGACTCCTATGAGCCCCATAAGGGTTCTGTGCTTAGCAATTAACAAAAAAGGTTTATA 1043
 Db 436 GCTGACTCCTATGAG-CCCAATAAGGGTTCTGTGCTTAGCAATTAACAAAAAAGGTTTATA 378
 QY 1044 GGTAAAGCCCAATGATTAATTTTGTGATGAGGCTTTAAAAATTTGCTCTCTTTT 1103
 Db 377 GGTAAAGCCCAATGATTAATTTTGTGATGAGGCTTTAAAAATTTGCTCTCTTTT 318
 QY 1104 CATATTTTATTCATATTCATATTTATGTTTGTAACTCTTTTAGGGAGATAATATATG 1163
 Db 317 CATATTTTATTCATATTCATATTTATGTTTGTAACTCTTTTAGGGAGATAATATATG 258
 QY 1164 TTATAAATTAGTTTGGGGGAATAATTTGTCAAGAGGATAATTTAATTTACGTGCTTC 1223
 Db 257 TTATAAATTAGTTTGGGGGAATAATTTGTCAAGAGGATAATTTAATTTACGTGCTTC 198
 QY 1224 TGTATTTCAGAAATAAGAGAGAGAGACTACGCTGCATATTCAGAGTTGTACCTTAACATT 1293
 Db 197 TGTATTTCAGAAATAAGAGAGAGAGACTACGCTGCATATTCAGAGTTGTACCTTAACATT 138
 QY 1284 GGTGAACATTTTCTTAAGATTTTCAAAAGGATAATGTGTAATTCAGAAATCATAAACC 1343
 Db 137 GGTGAACATTTTCTTAAGATTTTCAAAAGGATAATGTGTAATTCAGAAATCATAAACC 78
 QY 1344 ACTGTCCTAACTTGGTAAACAAACTGTTCTTTAAATAAGATTTAATGATTTTAAAAAAA 1403
 Db 77 ACTGTCCTAACTTGGTAAACAAACTGTTCTTTAAATAAGATTTAATGATTTTAAAAAAA 18
 QY 1404 AAAAAAAA 1413

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -40M13 fwd. from Amersham

High quality sequence stop: 366.

FEATURES

source

1. 444
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3804672"
/db_xref="taxon:9606"
/clone="IMAGE:490085"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Soares pregnant uterus NhPUP"
/note="Organ: uterus; Vector: pT773-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo (dt) primer [5',
AAGTGAAGATTCCGCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT773 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

ORIGIN

Query Match 31.0%; Score 444; DB 9; Length 444;
Best Local Similarity 100.0%; Pred. No. 1.2e-48;
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 963 AAGACTCATCTGTAGCTGTTGCTGACCTCTATGAGCCCATAGGGTCTGTGCTTACG 1022
Db 444 AAGACTCATCTGTAGCTGTTGCTGACCTCTATGAGCCCATAGGGTCTGTGCTTACG 385

QY 1023 ATTAACAAAATPAAGGTTTATAGGTAAGCCCAATGTAATTAATTTTGTGATGGAGGC 1082
Db 384 ATTAACAAAATPAAGGTTTATAGGTAAGCCCAATGTAATTAATTTTGTGATGGAGGC 325

QY 1083 TTTAAATTTGTCCTTTTTCATATTTTATTCATATTTCAATTTATGTTTGTAACTGCT 1142
Db 324 TTTAAATTTGTCCTTTTTCATATTTTATTCATATTTCAATTTATGTTTGTAACTGCT 265

QY 1143 TTTTAGGAGATAATTAATGCTTATAAATAGTTTGTGGGGGAATAATTTGCAAGAGG 1202
Db 264 TTTTAGGAGATAATTAATGCTTATAAATAGTTTGTGGGGGAATAATTTGCAAGAGG 205

QY 1203 ATAATTTAATTTACGCTCTCTGTTATTCAGAAATAAGAGAGAGACTACGCTGCATATT 1262
Db 204 ATAATTTAATTTACGCTCTCTGTTATTCAGAAATAAGAGAGAGACTACGCTGCATATT 145

QY 1263 CAAGAGTTGTACCTTAACATTTGTGAACATTTTTCATAGATTTTCAAAGGAATATGT 1322
Db 144 CAAGAGTTGTACCTTAACATTTGTGAACATTTTTCATAGATTTTCAAAGGAATATGT 85

QY 1323 GTAAATTCAGAAATCAATAACACCTGCTCTAACTTGGTAAACAAACTGTTCTTAATAAAG 1382
Db 84 GTAAATTCAGAAATCAATAACACCTGCTCTAACTTGGTAAACAAACTGTTCTTAATAAAG 25

QY 1383 TATTAAATGATTTTAAAAAATAA 1406
Db 24 TATTAAATGATTTTAAAAAATAA 1

RESULT 14

AW304923/c 438 bp mRNA linear EST 20-JAN-2000
LOCUS xv97b07.x1 NCI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2826421 3',
DEFINITION mRNA sequence.
ACCESSION AW304923
VERSION AW304923.1 GI:67117126
KEYWORDS EST.

SOURCE

ORGANISM

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1. (bases 1 to 438)

AUTHORS

NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index

JOURNAL

Unpublished (1998)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life

Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The

I.M.A.G.E. Consortium DNA Sequencing by: Washington University

Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyt not found

Seq primer: -40UP from Gibco

High quality sequence stop: 410.

FEATURES

source

1. 438

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2826421"

/tissue_type="three pooled meningiomas"

/lab_host="DH10B"

/clone_lib="NCI_CGAP_Brn53"

/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: SalI;

Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Library constructed by Life Technologies."

ORIGIN

Query Match 29.8%; Score 427; DB 10; Length 438;
Best Local Similarity 99.8%; Pred. No. 2e-46;
Matches 438; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 961 AGAAGACTCATCTGTAGCTGTTGCTGACCTCTATGAGCCCATAGGGTCTGTGCTTA 1020
Db 438 AGAAGACTCATCTGTAGCTGTTGCTGACCTCTATGAG-CCCATAGGGTCTGTGCTTA 380

QY 1021 GCATTAACAAAATAAGGTTTATAGGTAAGCCCAATGTAATTAATTTTTCATGGAGG 1080
Db 379 GCATTAACAAAATAAGGTTTATAGGTAAGCCCAATGTAATTAATTTTTCATGGAGG 320

QY 1081 GCATTAAATTTGTGCTCTTTTTCATATTTTATTCATATTTCAATTTATGTTTGTAACTG 1140
Db 319 GCATTAAATTTGTGCTCTTTTTCATATTTTATTCATATTTCAATTTATGTTTGTAACTG 260

QY 1141 CTTTTATAGGAGATAATTAATGTTATAAATTTAGTTTGGGGGAATAATTTGCAAGA 1200
Db 259 CTTTTATAGGAGATAATTAATGTTATAAATTTAGTTTGGGGGAATAATTTGCAAGA 200

QY 1201 GGAATAATTTTATGCTGCTTCTGTTTATTCAGATAAAGAGAGACTACGCTGCATA 1260
Db 199 GGAATAATTTTATGCTGCTTCTGTTTATTCAGATAAAGAGAGACTACGCTGCATA 140

QY 1261 TTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTTCATAGATTTTCAAAAGGAATAT 1320
Db 139 TTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTTCATAGATTTTCAAAAGGAATAT 80

QY 1321 GTGTAAATTCAGAAATCAATAACCACTGCTCTAACTTGGTAAACAAACTGTTCTTAATAA 1380
Db 79 GTGTAAATTCAGAAATCAATAACCACTGCTCTAACTTGGTAAACAAACTGTTCTTAATAA 20

QY 1381 AGTATTTAATGATTTTAAA 1399
Db 19 AGTATTTAATGATTTTAAA 1

Db 269 TTAGGAGATTAATAATGTTTAAATTAAGTTTGGGGGATTAATTTGCAAGAGGAT 210
QY 1205 AATTTAAATTTAGTGCTTTCTGTATTTCAGATAAAGAGAGAGAGACTAGCTGCATATTCA 1264
Db 209 AATTTAAATTTAGTGCTTTCTGTATTTCAGATAAAGAGAGAGAGACTAGCTGCATATTCA 150
QY 1265 AGAGTTGTACCTTAACATTTGTTGAAACATTTTCTTAAGATTTTCAAAAGGAATATGTGT 1324
Db 149 AGAGTTGTACCTTAACATTTGTTGAAACATTTTCTTAAGATTTTCAAAAGGAATATGTGT 90
QY 1325 AAATTGAGAAATCATAACCACTGTCTTAACCTTGTAAACAAACAACTGTCTTAATAAAGTA 1384
Db 89 AAATTGAGAAATCATAACCACTGTCTTAACCTTGTAAACAAACAACTGTCTTAATAAAGTA 30
QY 1385 TTTAATGATTTTAAAAAAGAAAAA 1413
Db 29 TTTGATGATTTTAAAAAAGAAAAA 1

Search completed: July 20, 2004, 19:51:20
Job time : 3474 secs

RESULT 15
AA491000/c
LOCUS
DEFINITION
aa52g11.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824612 3',
mRNA sequence.
ACCESSION
AA491000
VERSION
AA491000.1 GI:2220173
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 448)
REFERENCE
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -41m13 fwd. ET from Amersham
High quality sequence stop: 436.
FEATURES
Location/Qualifiers
1..448
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:824612"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GCB1"
/notes="vector: pT7T3D-Fac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IGD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGGCGGCGCTCATTTTCTTTTCTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 29.7%; Score 425.8; DB 9; Length 448;
Best Local Similarity 98.2%; Pred. No. 2.8e-46;
Matches 441; Conservative 0; Mismatches 7; Indels 1; Gaps 1;
QY 965 GACTCATCTGTAGCTGTTTGTGCTACTCTATGAGCCCATAGGGTCTGTGCTTAGCAT 1024
Db 448 GACTCATCTGTAGCTGTTTGTGCTACTCTATGAG-CCCATAGGGTCTGTGCTTAGCAT 390
QY 1025 TAACAAAATAGGTTTATAGGTAAAGCCAAATGATTAATTTTTCATGAGGGGCTT 1084
Db 389 TAACAAAATAGGTTTATAGGTAAAGCCAAATGATTAATTTTTCATGAGGGGCTT 330
QY 1085 TAAATTTTGTCTCTTTTTCATATTTTATCATATTCAATTTATGCTTTGTAACGCTTT 1144
Db 329 TAAATTTTGTCTCTTTTTCATATTTTATCATATTCATTTATGCTTTGTAACGCTTT 270
QY 1145 TTAGGAGATTAATTAATGTTTATAAATTAGTTTTCGGGGGAATAATTTGTCAAGAGGAT 1204

This Page Blank (uspto)

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 19:51:29 ; Search time 5499 Seconds
(without alignments)
11302.766 Million cell updates/sec

Title: US-10-799-747-12
Perfect score: 1434
Sequence: 1 cattaaactcttttctcg.....aaaaaaaaaaaaaaaaaaaa 1434

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.man.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1434	100.0	1434	6	BD078422	BD078422 101 human
2	735	51.3	3522	9	HSM804706	AL833393 Homo sapi
3	700	48.8	175081	9	AC021151	AC021151 Homo sapi
4	695	48.5	2020	9	HSM800227	AL049442 Homo sapi
5	631	44.0	3452	9	BC033650	BC033650 Homo sapi
6	400	27.9	281662	2	AC117374	AC117374 Homo sapi
7	345	24.1	448	6	AX408738	AX408738 Sequence
8	187	13.0	281662	2	AC117374	AC117374 Homo sapi
9	142	9.9	395	6	AX898296	AX898296 Sequence
10	142	9.9	395	6	BD033829	BD033829 Sequence
11	47	3.3	1648	10	BC063167	BC063167 Rattus no
12	47	3.3	8281	9	HSM801490	AL133623 Homo sapi
13	46	3.2	1378	3	AK113242	AK113242 Ciona int
14	46	3.2	110000	2	PFMAL13.16	Continuation (17 o
15	45	3.1	1648	9	BC009267	BC009267 Homo sapi
16	45	3.1	1772	10	BC058245	BC058245 Mus muscu
17	45	3.1	1922	9	AF232009	AF232009 Homo sapi
18	45	3.1	2156	9	AK025435	AK025435 Homo sapi
19	45	3.1	2240	9	BC036449	BC036449 Homo sapi
20	45	3.1	7084	9	HSM803370	AL832063 Homo sapi
21	44	3.1	1353	3	AY070501	AY070501 Drosophil
22	44	3.1	1452	3	AY058474	AY058474 Drosophil
23	44	3.1	1730	6	E52152	E52152 Gene encodi
24	44	3.1	1730	6	AR361006	AR361006 Sequence
25	44	3.1	1990	10	BC042668	BC042668 Mus muscu
26	44	3.1	2058	9	HSM802809	AL390184 Homo sapi
27	44	3.1	2515	10	BC006577	BC006577 Mus muscu
28	44	3.1	2539	5	BC045290	BC045290 Danio rer
29	44	3.1	2899	9	BC028198	BC028198 Homo sapi
30	44	3.1	3128	9	HSM807508	BX647364 Homo sapi
31	44	3.1	3487	10	BC016095	BC016095 Mus muscu
32	44	3.1	4075	9	BC021300	BC021300 Homo sapi
33	44	3.1	129517	2	AC016400	AC016400 Homo sapi
34	44	3.1	130503	2	AC130694	AC130694 Mus muscu
35	44	3.1	139756	2	AC144592	AC144592 Medicago
36	44	3.1	157274	2	AC140724	AC140724 Mus muscu
37	44	3.1	195221	10	AL683878	AL683878 Mouse DNA
38	44	3.1	211485	2	AC131304	AC131304 Mus muscu
39	44	3.1	213308	2	AC079164	AC079164 Mus muscu
40	44	3.1	237290	2	AC109543	AC109543 Rattus no
41	43	3.0	222	6	AR122013	AR122013 Sequence
42	43	3.0	384	6	AX185690	AX185690 Sequence
43	43	3.0	399	6	AR085693	AR085693 Sequence
44	43	3.0	399	6	I18358	I18358 Sequence 13
45	43	3.0	399	6	I21345	I21345 Sequence 13

ALIGNMENTS

RESULT 1	BD078422	101 human	secretory proteins.	1434 bp	DNA	linear	PAT 27-AUG-2002
LOCUS	BD078422	101 human	secretory proteins.				
DEFINITION	BD078422	101 human	secretory proteins.				
ACCESSION	BD078422	101 human	secretory proteins.				
VERSION	BD078422.1	GI:22624025					
KEYWORDS	JP 2001519156-A/11.						
SOURCE	Homo sapiens (human)						
ORGANISM	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.						
REFERENCE	1 (bases 1 to 1434)						
AUTHORS	Duan,R.D., Florence,K.A., Rosen,C.A., Ruben,S.M., Greene,J.M.,						
	Young,P., Ferrie,A.M., Yu,G.L., Janat,F., Ni,J., Carter,K.C.,						
	Endress,G.A., Feng,P., Lafleur,D.W. and Shi,Y.						

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Rutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3522)
Ottewaelder, B., Obermaier, B., Mewes, H.W., Weil, B. and Wiemann, S.
Direct Submission
Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by MediGenomix (Martinried/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp762K109) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clonesrzd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.
Location/Qualifiers
source
1. .3522
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp762K109"
/db_xref="taxon:9606"
/clone="DKFZp762K109"
/tissue type="melanoma (MeWo cell line)"
/clone_lib="762 (synonym: hm12). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"
polyA_signal
3420..3425
polyA_site
3440
ORIGIN

Query Match 51.3%; Score 735; DB 9; Length 3522;
Best Local Similarity 59.5%; Pred. No. 0;
Matches 1105; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

324 GTGGTAAATACAGAAATTTTACAGGCATTACCAAGCTAGGCATATAGCAATAGCAATGCGC 383
|||||
3269 GTGGTAAATACAGAAATTTTACAGGCATTACCAAGCTAGGCATATAGCAATAGCAATGCGC 2428
|||||

384 ACTCAGAAATGTTTCAATGTAGTAGTTGATGCTTTAGGTAGGGAGCTTATTCAGACA 443
|||||
2429 ACTCAGAAATGTTTCAATGTAGTAGTTGATGCTTTAGGTAGGGAGCTTATTCAGACA 2488
|||||

444 TAGTAGATGTTTCTCAATGCTGTSTCAATGCTGCGCTTTGGCTACCTGATCTCCSC 503
|||||
2489 TAGTAGATGTTTCTCAATGCTGTSTCAATGCTGCGCTTTGGCTACCTGATCTCCSC 2548
|||||

504 ATTATGCGACGCCATTGAGTCTTGAGTTTCTTCTCGACACCTTATGCTCTGAAATCA 563
|||||
2549 ATTATGCGACGCCATTGAGTCTTGAGTTTCTTCTCGACACCTTATGCTCTGAAATCA 2608
|||||

564 TGAGCGAGGCTGATTCATTTGGTGAATTTGGGTAGAAAGCAGTAGTTTTCCTGACATTAA 623
|||||
2609 TGAGCGAGGCTGATTCATTTGGTGAATTTGGGTAGAAAGCAGTAGTTTTCCTGACATTAA 2668
|||||

624 GATGAGGTTATAGATAGTTTACGCTTTAAGTGTATGTTTATATCTTTAAATAAGAA 683
|||||
2669 GATGAGGTTATAGATAGTTTACGCTTTAAGTGTATGTTTATATCTTTAAATAAGAA 2728
|||||

684 ATATAACCTTTAAGCTATTCCACCTCTCCCGAGCTATCTCAAACTGCTGGAATATA 743
|||||
2729 ATATAACCTTTAAGCTATTCCACCTCTCCCGAGCTATCTCAAACTGCTGGAATATA 2788
|||||

744 TGAGGAGATCTTGAAGAAGTAAATAAATCACTTCACTGCTCCACTCCAGGTGAATCGCC 803
|||||
2789 TGAGGAGATCTTGAAGAAGTAAATAAATCACTTCACTGCTCCACTCCAGGTGAATCGCC 2848
|||||

804 CACTCCCACTGACCTAGTAGAATTTGTAATTAATCTACCTCTTATTTCTGAATCAG 863
|||||
2849 CACTCCCACTGACCTAGTAGAATTTGTAATTAATCTACCTCTTATTTCTGAATCAG 2908
|||||

864 TTGTGAACCTGTGCTTATGTTTCAGARGTTTAAAGAACCTCMGTGAATTCATTTTAAAA 923
|||||
2909 TTGTGAACCTGTGCTTATGTTTCAGARGTTTAAAGAACCTCAGTGAATTCATTTTAAAA 2968
|||||

924 TCTGCTATTCTGAGAGCATTGAATGAATCTTAAACAAGAGACTCATCTGTAGCTGTTT 983
|||||
2969 TCTGCTATTCTGAGAGCATTGAATGAATCTTAAACAAGAGACTCATCTGTAGCTGTTT 3028
|||||

984 GCTGACTCCTATGAGCCCATATAAGGGTCTGTGCTTAGCATTAACAAAAAAGGTTTATA 1043
|||||
3029 GCTGACTCCTATGAG-CCCAATAAGGGTCTGTGCTTAGCATTAACAAAAAAGGTTTATA 3087
|||||

1044 GGTAAAGCCCAATGATTAAATTTTTCATGAGGGGCTTTAAAAATTTGTGCTCTTTT 1103
|||||
3088 GGTAAAGCCCAATGATTAAATTTTTCATGAGGGGCTTTAAAAATTTGTGCTCTTTT 3147
|||||

1104 CATATTTTATTCATATTCAATTTATGTTTGTAACTCTTTTAGGGAGATAAATATATG 1163
|||||
3148 CATATTTTATTCATATTCAATTTATGTTTGTAACTCTTTTAGGGAGATAAATATATG 3207
|||||

1164 TTATAAATTAGTTTGGGGGGAATAATTGTGCAAGAGGATAAATTTAATTTAGTGTCTTC 1223
|||||
3208 TTATAAATTAGTTTGGGGGGAATAATTGTGCAAGAGGATAAATTTAATTTAGTGTCTTC 3267
|||||

1224 TGTATTTCAGAAATAAAGAGAGAGACTACGCTGCATATTCAGAGTGTACCTTAACATT 1283
|||||
3268 TGTATTTCAGAAATAAAGAGAGAGACTACGCTGCATATTCAGAGTGTACCTTAACATT 3327
|||||

1284 GGTGAACCAATTTTCTAAGATTTTCAAAAGGAATATGTGTAATTTGAGAAATCATAACC 1343
|||||
3328 GGTGAACCAATTTTCTAAGATTTTCAAAAGGAATATGTGTAATTTGAGAAATCATAACC 3387
|||||

1344 ACTGCTCTAATCTGTTAAACAAACTGTTCTTAAATAAGATTAAATTTAATTTAAAAAAA 1403
|||||
3388 ACTGCTCTAATCTGTTAAACAAACTGTTCTTAAATAAGATTAAATTTAATTTAAAAAAA 3447
|||||

1404 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
|||||
3448 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3478
|||||

RESULT 3
AC021151/c 175081 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RFL1-483A20 from 4, complete sequence.
DEFINITION AC021151
ACCESSION AC021151
VERSION AC021151.8 GI:15145598
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Rutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 175081)
AUTHORS Kyung, K. and Abbott, A.
TITLE The sequence of Homo sapiens BAC clone RFL1-483A20
JOURNAL Unpublished (2001)
MEDLINE 3 (bases 1 to 175081)
PUBMED Waterston, R.H.
9847074
2 (bases 1 to 175081)
AUTHORS Kyung, K. and Abbott, A.
TITLE The sequence of Homo sapiens BAC clone RFL1-483A20
JOURNAL Unpublished (2001)
MEDLINE 3 (bases 1 to 175081)
PUBMED Waterston, R.H.
9847074
Direct Submission
Submitted (14-JAN-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 175081)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 175081)
AUTHORS Waterston, R.

TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Aug 9, 2001 this sequence version replaced gi:13877272.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapient@watson.wustl.edu
 ----- Summary Statistics

 Center project name: H_NH0483A20

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.regen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)
 VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the right is RP11-36G9. Actual start of this clone is at base position 1 of RP11-483A20; actual end is at base position 15081 of RP11-483A20.

FEATURES	source	Location/Qualifiers
	1..175081	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/chromosome="4"	
	/map="4"	
	/clone="RP11-483A20"	
	/clone_lib="RPCI-11"	
repeat_region	1..3267	
	/rpt_family="L1"	
repeat_region	3270..3453	
	/rpt_family="Alu"	
repeat_region	3434..3464	
	/rpt_family=" (A)n"	
repeat_region	3455..4161	
	/rpt_family="L1"	
misc_feature	7949..7956	
	/note="match to EST AW006057 (NID:G5954835) wz81d04.x1"	
repeat_region	10197..10489	
	/rpt_family="L1"	
repeat_region	10480..10501	
	/rpt_family="AT-rich"	
repeat_region	10495..10665	
	/rpt_family="L1"	
repeat_region	10737..11051	
	/rpt_family="L1"	
repeat_region	11052..11576	
	/rpt_family="L1"	
repeat_region	11564..11598	
	/rpt_family=" (A)n"	
repeat_region	11642..12324	
	/rpt_family="L1"	
repeat_region	11769..11794	
	/rpt_family="AT-rich"	
repeat_region	11939..11962	
	/rpt_family="AT-rich"	
repeat_region	12335..12412	
	/rpt_family="L1"	
repeat_region	12392..12621	
	/rpt_family="L1"	
repeat_region	12788..13096	
	/rpt_family="Alu"	
repeat_region	13069..13102	
	/rpt_family=" (CAAAA)n"	
repeat_region	13325..13667	
	/rpt_family="Alu"	
repeat_region	13444..13484	
	/rpt_family=" (TAAA)n"	
repeat_region	13640..13686	
	/rpt_family=" (GAAAA)n"	
repeat_region	14249..14422	
	/rpt_family="MER1_type"	
repeat_region	14980..15182	
	/rpt_family="L1"	
repeat_region	15260..15482	
	/rpt_family="L1"	
repeat_region	15483..15985	
	/rpt_family="MER1_type"	
repeat_region	15986..16105	
	/rpt_family="L1"	
repeat_region	16117..16251	
	/rpt_family="L1"	
repeat_region	16288..16356	
	/rpt_family="L1"	
repeat_region	16373..16545	
	/rpt_family="L1"	
repeat_region	16550..18773	
	/rpt_family="L1"	
repeat_region	18770..20218	
	/rpt_family="L1"	
repeat_region	19228..19346	
	/rpt_family="T-rich"	
repeat_region	20219..20514	
	/rpt_family="Alu"	
repeat_region	20515..22484	
	/rpt_family="L1"	
repeat_region	22584..22882	
	/rpt_family="Alu"	
repeat_region	22748..22768	
	/rpt_family="AT-rich"	
repeat_region	22883..23460	
	/rpt_family="L1"	
repeat_region	23461..23747	
	/rpt_family="MALR"	
repeat_region	23749..24514	
	/rpt_family="L1"	
misc_feature	24456	
	/note="match to EST AW902102 (NID:G8066307)"	
misc_feature	24488..24797	
	/note="match to EST AW902102 (NID:G8066307)"	
repeat_region	24576..24625	
	/rpt_family="Mariner"	
repeat_region	25484..25787	
	/rpt_family="Alu"	
repeat_region	25758..25787	
	/rpt_family=" (A)n"	
repeat_region	25796..26014	
	/rpt_family="MIR"	
repeat_region	26657..26889	

Qy	1139	TGCTTTTACGGAGATAATTATATGTTATTAATTTAGTTTGGGGGGAATAATTTGTGCAAA	1199
Db	1746	TGCTTTTACGGAGATAATTATATGTTATTAATTTAGTTTGGGGGGAATAATTTGTGCAAA	1805
Qy	1199	GAGGATAAATTTAATTTACGTCCTTCTGTTTATTCAGAAATAAGAGAGAAGACTACGCTGCA	1258
Db	1806	GAGGATAAATTTAATTTACGTCCTTCTGTTTATTCAGAAATAAGAGAGAAGACTACGCTGCA	1865
Qy	1259	TATTCAGAGAGTCTGCTTAACTAATCGTGAACATTTTCTTAAGATTTTCAAAAGGAAT	1318
Db	1866	TATTCAGAGAGTCTGCTTAACTAATCGTGAACATTTTCTTAAGATTTTCAAAAGGAAT	1925
Qy	1319	ATGTGTAAATTTGAGAATCATACACCTGTCCTAACTTGTGTAAACAACATTTCTTTAAAT	1378
Db	1926	ATGTGTAAATTTGAGAATCATACACCTGTCCTAACTTGTGTAAACAACATTTCTTTAAAT	1985
Qy	1379	AAAGTATTTAATGATTTTAAAAAATAAAAAA	1413
Db	1986	AAAGTATTTAATGATTTTAAAAAATAAAAAA	2020
RESULT 5			
LOCUS	BC033650		
DEFINITION	Homo sapiens hypothetical protein FLJ14431, mRNA linear PRI 06-OCT-2003		
ACCESSION	BC033650	IMAGE:5574637, complete cds.	
VERSION	BC033650.1	GI:21707066	
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Mammalia; Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 3452)		
	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altshuler S.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Hopkins R.F., Zolan B., Buetow K.H., Schaefer C.F., Bhat N.K., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udwin T.B., Toshiyuki S., Abramson P., Prange C., Raja S.S., Loquellano N.A., Peters G.J., Aronson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Wernley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Hailey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.O., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smalls D.E., Schnerch A., Schein J.E., Jones S.J. and Marra M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	99 (26),	16899-16903 (2002)
MEDLINE	22388257		
PUBMED	12477932		
REFERENCE	2 (bases 1 to 3452)		
AUTHORS	Strausberg R.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 1A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgabs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) Sequencing Center (NISC), Gaithersburg, Maryland;		

Qy	61	TTTGGAGCTGACAGTATTTTGTGTAGCAGAGATTTAAATTTTATATGAAAGTCAGTGC	120
Db	667	TTTGGAGCTGACAGTATTTTGTGTAGCAGAGATTTAAATTTTATATGAAAGTCAGTGC	726
Qy	121	AAAAATTATGATAGGATATCTAATAAATACAAAGTAAATACAAAGTCAAAGCAGTGT	180
Db	727	AAAAATTATGATAGGATATCTAATAAATACAAAGTAAATACAAAGTCAAAGCAGTGT	786
Qy	181	CTAAATAAAAAATTCGGGTTCCTTTAAATAATTTTAAATTTATC-TTGAATAGTTTTC	239
Db	787	CTAAATAAAAAATTCGGGTTCCTTTAAATAATTTTAAATTTATC-TTGAATAGTTTTC	846
Qy	240	TTGATTAATCTCAGGATATCAGAAAGTCATTAAGTGTGAGTAAAGTTAGTATCATTA	299
Db	847	TTGATTAATCTCAGGATATCAGAAAGTCATTAAGTGTGAGTAAAGTTAGTATCATTA	906
Qy	300	ACAAATTTGCTATTAAATGCGMGA-GTGGTAAATATACAGAAATTTATCAGGCATTACCAAG	358
Db	907	ACAAATTTGCTATTAAATGCGMGA-GTGGTAAATATACAGAAATTTATCAGGCATTACCAAG	966
Qy	359	TCTAGGCACATATAGGAATTCGACACTCAGAAATGGTTTCAATGATAGTTGATGCTTG	418
Db	967	TCTAGGCACATATAGGAATTCGACACTCAGAAATGGTTTCAATGATAGTTGATGCTTG	1026
Qy	419	TAAAGTAGGGAGCTTATTCAGACATAGTATAGTATTTCTTAATGCTGTSTCAATTTGCT	478
Db	1027	TAAAGTAGGGAGCTTATTCAGACATAGTATAGTATTTCTTAATGCTGTSTCAATTTGCT	1086
Qy	479	GGCTTTGGCTACCTGATCTTCCCATATGCGAGCCCATTCAGTCTGTGATTTCTTCT	538
Db	1087	GGCTTTGGCTACCTGATCTTCCCATATGCGAGCCCATTCAGTCTGTGATTTCTTCT	1146
Qy	539	CTGGACACTTATGCTCTGAATCATGAGCGAGCTGATCAATTTGTTGATTTGGGTAGA	598
Db	1147	CTGGACACTTATGCTCTGAATCATGAGCGAGCTGATCAATTTGTTGATTTGGGTAGA	1206
Qy	599	AAGCAGTATGTTTTGCTGACATTAAGATGTAGTTATAGATAGTTTAGCCTTTAAGTGT	658
Db	1207	AAGCAGTATGTTTTGCTGACATTAAGATGTAGTTATAGATAGTTTAGCCTTTAAGTGT	1266
Qy	659	ATGTTTTTATCTTAAATTAAGAAATATACTTTTAAGCTATTCACCTCTCCCTCCCA	718
Db	1267	ATGTTTTTATCTTAAATTAAGAAATATACTTTTAAGCTATTCACCTCTCCCTCCCA	1326
Qy	719	GCCTATCTCAAACTGGTGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCA	778
Db	1327	GCCTATCTCAAACTGGTGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCA	1386
Qy	779	CTGCTCCACTCAGGTGAATCCGCCACTCCACTGACCTAGTATGAAATTTTAAAT	838
Db	1387	CTGCTCCACTCAGGTGAATCCGCCACTCCACTGACCTAGTATGAAATTTTAAAT	1446
Qy	839	ACTTACCTTCTAATTTCTGAATCAGTTGTGAACCTGTGCTTATGTTTCAAGAGTTTAA	898
Db	1447	ACTTACCTTCTAATTTCTGAATCAGTTGTGAACCTGTGCTTATGTTTCAAGAGTTTAA	1506
Qy	899	ACCTCMGTGAATTCATTTTTTAAATCTGCTATTCTGAGAGCAATGAATGAATCTTAA	958
Db	1507	ACCTCMGTGAATTCATTTTTTAAATCTGCTATTCTGAGAGCAATGAATGAATCTTAA	1566
Qy	959	CAAGAAGACTCATCTGTAGCTGTTTGTGCTGCTCTATGAGCCCATAGGGTCTGTGCT	1018
Db	1567	CAAGAAGACTCATCTGTAGCTGTTTGTGCTGCTCTATGAG-CCCATAGGGTCTGTGCT	1625
Qy	1019	TAGCATTTAAACAAAATAAGGTTTATAGGTAAGCCAAATGATTAATTTTTTTTGCATGGA	1078
Db	1626	TAGCATTTAAACAAAATAAGGTTTATAGGTAAGCCAAATGATTAATTTTTTTTGCATGGA	1685
Qy	1079	GGGCTTTAAATTTGTGCTCTTTTTCATATTTTATTCATATTCATATTAAGTTTGTAAAC	1138
Db	1686	GGGCTTTAAATTTGTGCTCTTTTTCATATTTTATTCATATTTATGTTTGTAAAC	1745

Web site: <http://www.nisc.nih.gov/>

Contact: nisc_mgc@hri.nih.gov
 Akhter, N., Ayelle, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
 Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S.,
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
 Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
 McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
 Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAK Plate: 69 Row: n Column: 19
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 14249445.

FEATURES

Location/Qualifiers
 1..3452
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:44889 IMAGE:5574637"
 /tissue_type="Duodenum, adenocarcinoma"
 /clone_lib="NIH_MGC_88"
 /lab_host="DH10B"
 /notes="Vector: pCMV-SPORT6"
 1..3452
 /gene="FLJ14431"
 /db_xref="LocusID:84869"
 159..872
 /codon_start=1
 /product="FLJ14431 protein"
 /protein_id="AAH33650.1"
 /db_xref="GI:21707067"
 /db_xref="LocusID:84869"
 /translation="MDKVCVFGSGRIGRAVLAQLMARVRLAVLARNLEGAKARAG
 DLGGDLAFSCDAKEDHVDQNTPEELEKHLGRVNFVLAAGINRDLVRYTKEDMVS
 QLNTNLGSLTCKAARMRTMIQQQGSIVNGSLKNGSGSVYASRGGVLGFSR
 ALAKEVARKIRVNVVAPGVFDMTKDLKBEHLKKNKPLIGREFGETIEVAHAVVFLE
 SPYITGHVLVDGGLQLL"
 159..857
 /notes="adh short; Region: short chain dehydrogenase. This
 family contains a wide variety of dehydrogenases"
 /db_xref="CDD:pfam00106"

ORIGIN

Query Match 44.0%; Score 631; DB 9; Length 3452;
 Best Local Similarity 99.3%; Pred. No. 6.1e-305;
 Matches 1291; Conservative 0; Mismatches 6; Indels 3; Gaps 3;
 1 CATTAACTCTTTTATCGGGAATAGTAGATATTTTCAATGTCACCTCCATTCATGTGA 60
 2034 CATTAACTCTTTTATCGGGAATAGTAGATATTTTCAATGTCACCTCCATTCATGTGA 2093
 61 TTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTATTTTATATTTGAAGTCAGTGC 120
 2094 TTGGAGCTGACAGTTATTTTGTGAAGCAGAGATTATTTTATATTTGAAGTCAGTGC 2153
 121 AAAATATTGAATAGGATATATTAATAAATAACAAGTAATAACAAAAGTCAAGCAGTGT 180
 2154 AAAATATTGAATAGGATATATTAATAAATAACAAGTAATAACAAAAGTCAAGCAGTGT 2213
 181 CTAATAAATAAATCTGGGTTCCCTTAAATAATTTTAAATTTATC-TTGAATAGTTTC 239
 2214 CTAATAAATAAATCTGGGTTCCCTTAAATAATTTTAAATTTATC-TTGAATAGTTTC 2273
 240 TTAGATTATCTCAGGATATGAGAAATCAATTAAGTGTGAGTAAGTTAGTATCAATTA 299
 2274 TTAGATTATCTCAGGATATGAGAAATCAATTAAGTGTGAGTAAGTTAGTATCAATTA 2333
 300 ACAAATGTCTATTAATGACGAGTGGTAAATATACAGAAATTTATCAGGCATTTACCAAG 358
 2334 ACAAATGTCTATTAATGACGAGTGGTAAATATACAGAAATTTATCAGGCATTTACCAAG 2393

QY 359 TCTAGGCACATATAGGAAATGCAGCACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTG 418
 DB 2394 TCTAGGCACATATAGGAAATGCAGCACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTG 2453
 QY 419 TAAGGTAGGGAGCTTATTTCAGACATAGTAGATAGTTTCTCTAATGTCTGTSTCAATGTCT 478
 DB 2454 TAAGGTAGGGAGCTTATTTCAGACATAGTAGATAGTTTCTCTAATGTCTGTCTCAATGTCT 2513
 QY 479 GGCTTTTGGCTACTCTGCTTCCSCATTAATGGCAGGCCATTTCACTGTCTTGTAGTTTCTTCT 538
 DB 2514 GGCTTTTGGCTACTCTGCTTCCSCATTAATGGCAGGCCATTTCACTGTCTTGTAGTTTCTTCT 2573
 QY 539 CTGACACACCTTATGCTCTGAAATCATGAGCAGGCTGATTCAATTTGGTGAATTTGGGTAGA 598
 DB 2574 CTGACACACCTTATGCTCTGAAATCATGAGCAGGCTGATTCAATTTGGTGAATTTGGGTAGA 2633
 QY 599 AGCAGTATGTTTTGCTGACATTAAGATGTAGGTATAGATAGTTTGTAGCTTTTAAAGTGT 658
 DB 2634 AGCAGTATGTTTTGCTGACATTAAGATGTAGGTATAGATAGTTTGTAGCTTTTAAAGTGT 2693
 QY 659 ATGTTTTTATACCTTTAAATATAAGAAATATAACCTTTTAAAGTATTCACCTCTCCCTCCA 718
 DB 2694 ATGTTTTTATACCTTTAAATATAAGAAATATAACCTTTTAAAGTATTCACCTCTCCCTCCA 2753
 QY 719 GCCTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCA 778
 DB 2754 GCCTATCTCAAACTGGTGGAAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCA 2813
 QY 779 CTGCTCCACTCCAGGTGAATCCGCCCACTCCCACTAGCTAGTAGAATTTGTATTTAAT 838
 DB 2814 CTGCTCCACTCCAGGTGAATCCGCCCACTCCCACTAGCTAGTAGAATTTGTATTTAAT 2873
 QY 839 ACTTACTCTCTATTTCTGAAATCAGTTGTGAACCTGTTGCCTTATGTTTCAGARGTTTAAAGA 898
 DB 2874 ACTTACTCTCTATTTCTGAAATCAGTTGTGAACCTGTTGCCTTATGTTTCAGARGTTTAAAGA 2933
 QY 899 ACCTCMGTGAATTCATTTTTTAAATCTGCTATTTCTGAGAAGCATTTGAATGAATTTCTTAA 958
 DB 2934 ACCTCAGTGAATTCATTTTTTAAATCTGCTATTTCTGAGAAGCATTTGAATGAATTTCTTAA 2993
 QY 959 CAAGAAGACTCATCTGCTAGTGTGCTGACTCTCTATGAGCCCACTAAGGGTTCGTGCT 1018
 DB 2994 CAAGAAGACTCATCTGCTAGTGTGCTGACTCTCTATGAG-CCCACTAAGGGTTCGTGCT 3052
 QY 1019 TAGCATTAACAAATAAAGTTTATAGGTAAGCAATGTATTAAATTTTTTTTGCATGGA 1078
 DB 3053 TAGCATTAACAAATAAAGTTTATAGGTAAGCAATGTATTAAATTTTTTTTGCATGGA 3112
 QY 1079 GGGCTTTAAATTTGTGCTCTTTTCAATTTTATTTCAATTTTCAATTTTATGTTTGTAAAC 1138
 DB 3113 GGGCTTTAAATTTGTGCTCTTTTCAATTTTATTTTCAATTTTCAATTTTATGTTTGTAAAC 3172
 QY 1139 TGGTTTTTGGGAGATATATATGTTATAAATAGTTTGGGGGAAATAATTTGTCAAA 1198
 DB 3173 TGGTTTTTGGGAGATATATATGTTATAAATAGTTTGGGGGAAATAATTTGTCAAA 3232
 QY 1199 GAGGATAATTTTAAATTTACGTGCTTCTGTTATTATTTAGGTAAGCAAGTACGCTGCA 1258
 DB 3233 GAGGATAATTTTAAATTTACGTGCTTCTGTTATTATTTAGGTAAGCAAGTACGCTGCA 3292
 QY 1259 TATTCAGAGTTGTACCTTTAAACATTTGGTGAACATTTTTT 1298
 DB 3293 TATTCAGAGTTGTACCTTTAAACATTTGGTGAACATTTTTT 3332
 RESULT 6
 AC117374 281662 bp DNA linear HTG 02-MAY-2002
 LOCUS Homo sapiens chromosome 12 clone RP11-19D19, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 48 unordered pieces.
 AC117374 AC011694
 AC117374.1 GI:20127352


```

* 102349 102448: gap of unknown length
* 102449 105973: contig of 3525 bp in length
* 105974 106073: gap of unknown length
* 106074 112891: contig of 6818 bp in length
* 112891 112991: gap of unknown length
* 112991 116221: contig of 3330 bp in length
* 116221 116321: gap of unknown length
* 116321 117998: contig of 3477 bp in length
* 117998 119899: gap of unknown length
* 119899 124870: contig of 4972 bp in length
* 124870 124970: gap of unknown length
* 124970 130790: contig of 5820 bp in length
* 130790 130890: gap of unknown length
* 130890 134639: contig of 3749 bp in length
* 134639 134739: gap of unknown length
* 134740 140637: contig of 5898 bp in length
* 140638 140737: gap of unknown length
* 140738 145541: contig of 4804 bp in length
* 145541 145641: gap of unknown length
* 145641 15057: contig of 9416 bp in length
* 15057 15157: gap of unknown length
* 15157 164357: contig of 9200 bp in length
* 164358 164457: gap of unknown length
* 164458 174892: contig of 10435 bp in length
* 174893 174992: gap of unknown length
* 174993 184400: contig of 9408 bp in length
* 184401 18500: gap of unknown length
* 18500 198385: contig of 13885 bp in length
* 198386 198485: gap of unknown length
* 198486 216436: contig of 17951 bp in length
* 216437 216536: gap of unknown length
* 216537 232916: contig of 16380 bp in length
* 232917 233016: gap of unknown length
* 233017 255437: contig of 22421 bp in length
* 255438 255537: gap of unknown length
* 255538 281662: contig of 26125 bp in length.
FEATURES
    source
        1..281662
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="12"
            /clone="RP11-19D19"
Query Match      27.9%; Score 400; DB 2; Length 281662;
Best Local Similarity 100.0%; Pred. No. 1.9e-189;
Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1000 CCNATAAGGTTCTGCTTAGCATTACAAAATAAGGTTTATAGGTAAGCCCAATGTAT 1059
Db 34088 CCNATAAGGTTCTGCTTAGCATTACAAAATAAGGTTTATAGGTAAGCCCAATGTAT 34147
QY 1060 TAAATTTTTTTTGCATGGAGGCTTTAAATTTCTGCTCTTTTCATATTTTATTCATAT 1119
Db 34148 TAAATTTTTTTTGCATGGAGGCTTTAAATTTCTGCTCTTTTCATATTTTATTCATAT 34207
QY 1120 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTATATGTTATAAATTAGTTTGT 1179
Db 34208 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTATATGTTATAAATTAGTTTGT 34267
QY 1180 GGGGGAATAATTGCAAGAGGATAATTAAATTTAGTGTCTTCTGTTTATTCAGAAATAA 1239
Db 34268 GGGGGAATAATTGCAAGAGGATAATTAAATTTAGTGTCTTCTGTTTATTCAGAAATAA 34327
QY 1240 GAGAGAAGACTACGCTGCATTTCAAGAGTTGTACCTTTAACTTTGAGGTAACATTTTTC 1299
Db 34328 GAGAGAAGACTACGCTGCATTTCAAGAGTTGTACCTTTAACTTTGAGGTAACATTTTTC 34387
QY 1300 TAAGATTTTCAAAGGAATATGTTAAATTCAGAAATCATAACCACTGTCTTAACCTTGGT 1359
Db 34388 TAAGATTTTCAAAGGAATATGTTAAATTCAGAAATCATAACCACTGTCTTAACCTTGGT 34447

```

```

QY 1360 AACAAACTGTTCTTAATAAAGTATTTAATGATTTTAAA 1399
Db 34448 AACAAACTGTTCTTAATAAAGTATTTAATGATTTTAAA 34487
RESULT 7
AX408738/c
LOCUS AX408738 448 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 1385 from Patent WO0229103.
ACCESSION AX408738
VERSION AX408738.1 GI:21441443
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 1385 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
    source
        1..448
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            /note="EMBL/GenBank Accession No. AA491000"
ORIGIN
Query Match      24.1%; Score 345; DB 6; Length 448;
Best Local Similarity 100.0%; Pred. No. 1.3e-161;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1000 CCCATAAGGTTCTGCTTAGCATTACAAAATAAGGTTTATAGGTAAGCCCAATGTAT 1059
Db 414 CCCATAAGGTTCTGCTTAGCATTACAAAATAAGGTTTATAGGTAAGCCCAATGTAT 355
QY 1060 TAAATTTTTTTTGCATGGAGGCTTTAAATTTCTGCTCTTTTCATATTTTATTCATAT 1119
Db 354 TAAATTTTTTTTGCATGGAGGCTTTAAATTTCTGCTCTTTTCATATTTTATTCATAT 295
QY 1120 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTATATGTTATAAATTAGTTTGT 1179
Db 294 TCAATTTATGTTTGTAACTGCTTTTATAGGAGATAATTATATGTTATAAATTAGTTTGT 235
QY 1180 GGGGGAATAATTGCAAGAGGATAATTAAATTTAGTGTCTTCTGTTTATTCAGAAATAA 1239
Db 234 GGGGGAATAATTGCAAGAGGATAATTAAATTTAGTGTCTTCTGTTTATTCAGAAATAA 175
QY 1240 GAGAGAAGACTACGCTGCATTTCAAGAGTTGTACCTTTAACTTTGAGGTAACATTTTTC 1299
Db 174 GAGAGAAGACTACGCTGCATTTCAAGAGTTGTACCTTTAACTTTGAGGTAACATTTTTC 115
QY 1300 TAAGATTTTCAAAGGAATATGTTAAATTCAGAAATCATAACCA 1344
Db 114 TAAGATTTTCAAAGGAATATGTTAAATTCAGAAATCATAACCA 70
RESULT 8
AC117374/c
LOCUS AC117374 281662 bp DNA linear HTG 02-MAY-2002
DEFINITION Homo sapiens chromosome 12 clone RP11-19D19, *** SEQUENCING IN
            PROGRESS ***, 48 unordered pieces.
ACCESSION AC117374 AC011694
VERSION AC117374.1 GI:20127352
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
            Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,

```


Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Garroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Halle, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lorado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mahwiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
 Oreguen, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 281662)
 Worley, K.C.

Submitted (10-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 281662)
 Worley, K.C.

Direct Submission
 Submitted (02-MAY-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On Apr 28, 2002 this sequence version replaced gi:10045385.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: HCNB
 Center clone name: RP11-19D19
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 273063 bases at least Q40
 Consensus quality: 294336 bases at least Q30
 Consensus quality: 306942 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 48 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 2464: contig of 2464 bp in length
 2465: gap of unknown length
 2564: contig of 2039 bp in length
 2565: gap of unknown length
 4603: contig of 2410 bp in length
 4703: gap of unknown length
 7113: contig of 2304 bp in length
 7214: gap of unknown length
 9518: contig of 2490 bp in length
 9618: gap of unknown length
 12107: contig of 2807 bp in length
 12208: gap of unknown length
 15014: contig of 2717 bp in length
 15114: gap of unknown length
 15135: contig of 2549 bp in length
 17832: gap of unknown length
 17932: contig of 2120 bp in length
 20481: gap of unknown length
 20581: contig of 2109 bp in length
 22701: gap of unknown length
 22801: contig of 2304 bp in length
 24910: gap of unknown length
 25010: contig of 2385 bp in length
 27314: gap of unknown length
 27414: contig of 3568 bp in length
 29799: gap of unknown length
 29899: contig of 3113 bp in length
 33567: gap of unknown length
 33679: contig of 3693 bp in length
 36680: gap of unknown length
 36780: contig of 2111 bp in length
 40473: gap of unknown length
 40573: contig of 3536 bp in length
 42684: gap of unknown length
 46320: contig of 3088 bp in length
 46420: gap of unknown length
 49508: contig of 3605 bp in length
 49608: gap of unknown length
 53213: contig of 3955 bp in length
 53313: gap of unknown length
 57268: contig of 3077 bp in length
 57368: gap of unknown length
 60445: contig of 5315 bp in length
 60545: gap of unknown length
 65860: contig of 3079 bp in length
 65960: gap of unknown length
 69039: contig of 4307 bp in length
 69139: gap of unknown length
 73446: contig of 4562 bp in length
 73546: gap of unknown length
 78108: contig of 4297 bp in length
 82505: gap of unknown length
 82605: contig of 4372 bp in length
 86977: gap of unknown length
 91316: contig of 4239 bp in length
 91416: gap of unknown length
 95085: contig of 3669 bp in length
 95185: gap of unknown length
 102349: contig of 7164 bp in length
 102449: gap of unknown length
 105973: contig of 3525 bp in length
 105974: gap of unknown length
 112891: contig of 6818 bp in length
 112892: gap of unknown length
 116221: contig of 3230 bp in length
 116222: gap of unknown length
 119798: contig of 3477 bp in length

```
* 119799 119898: gap of unknown length
* 119899 124870: contig of 4972 bp in length
* 124871 124970: gap of unknown length
* 124971 130790: contig of 5820 bp in length
* 130791 130890: gap of unknown length
* 130891 134639: contig of 3749 bp in length
* 134640 134739: gap of unknown length
* 134740 140637: contig of 5898 bp in length
* 140638 140738: gap of unknown length
* 140739 145541: contig of 4804 bp in length
* 145542 145641: gap of unknown length
* 145642 155057: contig of 9416 bp in length
* 155058 155157: gap of unknown length
* 155158 164357: contig of 9200 bp in length
* 164358 164457: gap of unknown length
* 164458 174892: contig of 10435 bp in length
* 174893 174993: gap of unknown length
* 174993 184400: contig of 9408 bp in length
* 184401 184500: gap of unknown length
* 184501 198385: contig of 13885 bp in length
* 198386 198485: gap of unknown length
* 198486 216436: contig of 17951 bp in length
* 216437 232916: gap of unknown length
* 232917 233017: contig of 16380 bp in length
* 233017 255438: gap of unknown length
* 255438 255537: gap of unknown length
* 255538 281662: contig of 26125 bp in length.
FEATURES
  source
    1. .281662
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /chromosome="12"
      /clone="RP11-19D19"
  ORIGIN
    Query Match      13.0%; Score 187; DB 2; Length 281662;
    Best Local Similarity 100.0%; Pred. No. 9.2e-83;
    Matches 187; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 41 TGTCACTCCATTCATGTTGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTAAAT 100
  Db 146186 TGTCACTCCATTCATGTTGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTAAAT 146127
  QY 101 TTTATATTGAAGTCAGTCAAAATTTATGAATAGGATATATCTATATAATACAAAGTAATA 160
  Db 146126 TTTATATTGAAGTCAGTCAAAATTTATGAATAGGATATATCTATATAATACAAAGTAATA 146067
  QY 161 ACAAAGTCGAAGCAGTGTCTTAATAATAAATTTCTGGTTCCCTTAAAAATTTTAAAT 220
  Db 146066 ACAAAGTCGAAGCAGTGTCTTAATAATAAATTTCTGGTTCCCTTAAAAATTTTAAAT 146007
  QY 221 TTATCTT 227
  Db 146006 TTATCTT 146000
  RESULT 9
  AX898296
  LOCUS
    AX898296 395 bp DNA linear PAT 18-DEC-2003
  DEFINITION
    Sequence 14159 from Patent EP1033401.
  ACCESSION
    AX898296
  VERSION
    AX898296.1 GI:40053209
  KEYWORDS
    Homo sapiens (human)
  SOURCE
    Homo sapiens
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    Dumas Milne Edwards,J.B., Duclert,A. and Giordano,J.Y.
    Expressed sequence tags and encoded human proteins
    Patent: EP 1033401-A 14159 06-SEP-2000;
  REFERENCE
    1
  AUTHORS
    Edwards,J.B.D.M., Duclert,A. and Jordan,J.Y.
  TITLE
    Sequence tag and encoded human protein
  JOURNAL
    Patent: JP 2001269182-A 10075 02-OCT-2001;
  GENSSET
    OS Homo sapiens (human)
    PN JP 2001269182-A/10075
    PD 02-OCT-2001
    PF 24-FEB-2000 JP 2000118773
    PR 26-FEB-1999 US 60/122487
    PI JEAN RAPUIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
    PI JORDAN
    PC C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21,PC
    C12N5/10,
    PC C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00,PC
    G06F15/40
  CC Key Location/Qualifiers.
  FH Key Location/Qualifiers
  1. .395
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
  ORIGIN
    Query Match      9.9%; Score 142; DB 6; Length 395;
    Best Local Similarity 100.0%; Pred. No. 5.9e-60;
    Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 58 TGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTATTTATTTGAAAGTCAG 117
  Db 1 TGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTATTTATTTGAAAGTCAG 60
  QY 118 TGC AAAATTTATGAATAGGATATCTATATAATACAAAGTAATAACAAAGTCAAAGCAGT 177
  Db 61 TGC AAAATTTATGAATAGGATATCTATATAATACAAAGTAATAACAAAGTCAAAGCAGT 120
  QY 178 GTTCTAAATAAAAATTCCTGGT 199
  Db 121 GTTCTAAATAAAAATTCCTGGT 142
  RESULT 10
  BD033829
  LOCUS
    BD033829 395 bp DNA linear PAT 27-AUG-2002
  DEFINITION
    Sequence tag and encoded human protein.
  ACCESSION
    BD033829
  VERSION
    BD033829.1 GI:22575571
  KEYWORDS
    JP 2001269182-A/10075.
  SOURCE
    Homo sapiens (human)
  ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    Edwards,J.B.D.M., Duclert,A. and Jordan,J.Y.
    Sequence tag and encoded human protein
    Patent: JP 2001269182-A 10075 02-OCT-2001;
  GENSSET
    OS Homo sapiens (human)
    PN JP 2001269182-A/10075
    PD 02-OCT-2001
    PF 24-FEB-2000 JP 2000118773
    PR 26-FEB-1999 US 60/122487
    PI JEAN RAPUIST DUMAS MILNE EDWARDS,EIMERIC DUCLAIR,JEAN YVES
    PI JORDAN
    PC C12N15/09,C07K14/435,C07K16/18,C12N1/15,C12N1/19,C12N1/21,PC
    C12N5/10,
    PC C12P21/02,C12P21/08,C12Q1/68//G06F17/30,C12N15/00,C12N5/00,PC
    G06F15/40
  CC Key Location/Qualifiers.
  FH Key Location/Qualifiers
  1. .395
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
  ORIGIN
    Query Match      9.9%; Score 142; DB 6; Length 395;
    Best Local Similarity 100.0%; Pred. No. 5.9e-60;
    Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 58 TGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTATTTATTTGAAAGTCAG 117
  Db 1 TGATTTGGAGCTGACAGTATTATTTGCTGAAGCAGAGATTATTTATTTGAAAGTCAG 60
  QY 118 TGC AAAATTTATGAATAGGATATCTATATAATACAAAGTAATAACAAAGTCAAAGCAGT 177
  Db 61 TGC AAAATTTATGAATAGGATATCTATATAATACAAAGTAATAACAAAGTCAAAGCAGT 120
```



```

QY 178 GTTCTAATAAAAAATTCGGGT 199
Db 121 GTTCTAATAAAAAATTCGGGT 142

RESULT 11
BC063167 1648 bp mRNA linear ROD 03-DEC-2003
LOCUS Rattus norvegicus ADP-ribosylation factor 4, mRNA (cDNA clone
MGC:72808 IMAGE:6921397), complete cds.
ACCESSION BC063167
VERSION BC063167.1 GI:38648903
KEYWORDS MGC.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1648)
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buettow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullighy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalish, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 1648)
Straussberg, R.
Direct Submission
Submitted (02-DEC-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: John C. Marshall, M.D., Ph.D
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
BCA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: IRAL Plate: 52 Row: 0 Column: 8
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 13162342.
Location/Qualifiers
1. .1648

/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="MGC:72808 IMAGE:6921397"
/tissue_type="Pituitary gland, anterior, rat"
/clone_lib="NICHDR_Pit1"
/lab_host="DH10B"
/notes="vector: pDNR-LIB"
1. .1648
/gene="Arf4"
/db_xref="LocusID:79120"
/db_xref="RGD:621275"
178. .720
/codon_start=1
/product="ADP-ribosylation factor 4"
/protein_id="AAH63167.1"
/db_xref="GI:38648904"
/db_xref="LocusID:79120"
/translation="MGLTISLSFRLPKQKQIRLMLVGLDAAGKTYLKLGLIEIVT
TPTIGNVEIVKNTCFVWDVGGQDKIRPLWRHYFQNTQGLIFVVDSDNRERIOE
GAAVLQWLLLEDELODAVLLLEFANKQDLPNMAITSEMTDKLGQLSLNRRTWYVQATCA
TQGTGLYEGLDMLSNLSKR"
178. .717
/notes="ARF; Region: ARF-like small GTPases"
/db_xref="CDD:smart00177"

ORIGIN
Query Match 3.3%; Score 47; DB 10; Length 1648;
Best Local Similarity 100.0%; Pred. No. 2e-12;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AATGATTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
|||||
Db 1556 AATGATTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1602
|||||

RESULT 12
HSM801490 8281 bp mRNA linear PRI 18-FEB-2000
LOCUS Homo sapiens mRNA; cDNA DKFZp434P0721 (from clone DKFZp434P0721);
partial cds.
ACCESSION AL133623
VERSION AL133623.1 GI:6599261
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 8281)
Blum, H., Bauersachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
Direct Submission
Submitted (15-DEC-1999) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp434P0721) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
Location/Qualifiers
1. .8281
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="3"
/clone="DKFZp434P0721"
/tissue_type="testis"
/clone_lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"

FEATURES
source

```

```

Fax:81-75-705-1113)
Ciona intestinalis cDNA Project (URL:
http://ghost.zool.kyoto-u.ac.jp/index1.html).
FEATURES
    source                1..1378
                        /organism="Ciona intestinalis"
                        /mol_type="mRNA"
                        /db_xref="taxon:7719"
                        /clone="ciad019cl3"
ORIGIN
Query Match          3.2%; Score 46; DB 3; Length 1378;
Best Local Similarity 100.0%; Pred. No. 6,3e-12;
Matches 46; Conservative 0; Mismatches 0; Indels 0;

Qy 1389 ATGATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
      |||
Db 1328 ATGATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1373

RESULT 14
PFMAL13_16
WPCOMMENT
Sequence split into 28 fragments LOCUS PFMAL13 Accession AL844509
Fragment Name      Begin      End
PFMAL13_00         1      110000
PFMAL13_01         100001   210000
PFMAL13_02         200001   310000
PFMAL13_03         300001   410000
PFMAL13_04         400001   510000
PFMAL13_05         500001   610000
PFMAL13_06         600001   710000
PFMAL13_07         700001   810000
PFMAL13_08         800001   910000
PFMAL13_09         900001  1010000
PFMAL13_10        1000001  1110000

```

Query Match	Best Local Similarity	3.2%;	Score 46;	DB 2;	Length 110000;
Matches 46;	Conservative	100.0%;	Pred. No. 4.1e-12;	Mismatches 0;	Indels 0;

Qy	1389	ATGATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	1434
Db	20114	ATGATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	20159

RESULT 15			
BC009267	LOCUS	BC009267	1648 bp
	DEFINITION	Homo sapiens lamin B2, mRNA (cDNA clone IMAGE:4053047), cds.	linear
	ACCESSION	BC009267	
	VERSION	BC009267.2	GI:33872541
	KEYWORDS	.	
	SOURCE	Homo sapiens (human)	

Search completed: July 20, 2004, 23:04:25
Job time : 5505 secs

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 1648)
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heien,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Sapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,D.J., Hulyk,S.W.,
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL 22388257
MEDLINE 12477932
REFERENCE 2 (bases 1 to 1648)
Strausberg,R.
Direct Submission
Submitted (08-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT On Aug 19, 2003 this sequence version replaced gi:14349378.
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbiology.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettman, Anuradha
Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: JRAL Plate: 25 Row: 0 Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 14249359.
FEATURES
source
location/Qualifiers
1..1648
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4033047"
/tissue_type="uterus, leiomyosarcoma"
/clone_lib="NIH MGC_46"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

ORIGIN
Query Match 3.1%; Score 45; DB 9; Length 1648;
Best Local Similarity 100.0%; Pred. No. 2e-11;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
|||||
DB 1604 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1648

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 18:53:19 ; Search time 583 Seconds
(without alignments)
10449.256 Million cell updates/sec

Title: US-10-799-747-12
Perfect score: 1434
Sequence: 1 cattaacttttttctcg.....aaaaaaaaaaaaaaaaaa 1434

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1434	100.0	1434	7	ADA39771 Human sec
2	1434	100.0	1434	7	ACC50424 Human sec
3	1383	96.4	1434	2	AAX37452 Human sec
C 4	345	24.1	448	6	ABN94887 Gene #138
5	145	10.1	415	3	AHH30357 Human col
6	142	9.9	395	3	AAC10084 Human sec
7	127	8.9	336	2	AAT25136 Human gen
C 8	45	3.1	493	9	ADE81840 Arabidops
9	45	3.1	1231	4	AHH25190 Nucleotid
10	45	3.1	1348	2	AZ224411 Human bla
11	45	3.1	2552	6	ABL89915 Human pol
12	45	3.1	3470	5	ABV25066 Human pro
13	44	3.1	1730	3	AAH87997 Torenla h
14	43	3.0	222	2	AAQ46071 Sequence
15	43	3.0	222	4	AAC86777 Downstrea
C 16	43	3.0	271	4	AAL16362 Human bre
17	43	3.0	306	6	ABZ08725 Human leu
C 18	43	3.0	337	5	ABV55706 Human pro
C 19	43	3.0	337	6	ABQ86019 Arabidops
C 20	43	3.0	339	4	AAI86180 Human pol
C 21	43	3.0	380	5	ABAI3116 Human ner
22	43	3.0	383	6	ABL01579 Human sec
C 23	43	3.0	384	4	AAH70111 Human cer

24	43	3.0	399	2	AAT13044	Aat13044 Cotton fi
25	43	3.0	399	2	AAT30261	Aat30261 Cotton fi
C 26	43	3.0	399	2	AAT62620	Aat62620 Cotton fi
27	43	3.0	399	2	AAT70051	Aat70051 Cotton fi
C 28	43	3.0	399	3	AZ35555	Az35555 cDNA sequ
29	43	3.0	401	4	AAI88370	Aai88370 Human pol
30	43	3.0	402	4	AAI85111	Aai85111 Human pol
31	43	3.0	402	8	ACH18461	Ach18461 Human adu
32	43	3.0	418	4	AAI88540	Aai88540 Human pol
33	43	3.0	429	5	ABV49095	Abv49095 Human pro
C 34	43	3.0	480	6	ABZ08201	Abz08201 Human leu
C 35	43	3.0	528	6	ABZ08195	Abz08195 Human leu
C 36	43	3.0	556	5	ABV40163	Abv40163 Human pro
C 37	43	3.0	556	5	ABV40063	Abv40063 Human pro
C 38	43	3.0	556	5	ABV42105	Abv42105 Human pro
C 39	43	3.0	556	5	ABV43601	Abv43601 Human pro
C 40	43	3.0	638	6	ABZ78098	Abz78098 Human bre
41	43	3.0	681	5	ABV19321	Abv19321 Human pro
42	43	3.0	785	3	AAC77913	Aac77913 Human can
43	43	3.0	916	2	AAT65651	Aat65651 cDNA enco
44	43	3.0	916	2	AAV62000	Aav62000 R. prolix
45	43	3.0	944	4	AAH32549	Aah32549 Human sec

ALIGNMENTS

RESULT 1					
ADA39771					
ID	ADA39771	standard;	cDNA;	1434	BP.
XX	ADA39771;				
XX	20-NOV-2003	(first entry)			
DT	XX				
DE	Human secreted protein encoding cDNA.				
XX	Human; secreted protein; cancer; hyperproliferative disorder;				
KW	rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;				
KW	anaemia; allergic reaction; asthma; cardiovascular disorder;				
KW	wound healing; cytostatic; immunosuppressive; neutropenic; neuroprotective;				
KW	antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;				
XX	vulnerary; cardiac; gene therapy; ss.				
OS	Homo sapiens.				
XX	WO2002102993-A2.				
PN	27-DEC-2002.				
XX	19-MAR-2002; 2002WO-US008123.				
PF	21-MAR-2001; 2001US-0277340P.				
PR	19-JUL-2001; 2001US-0306171P.				
PR	13-NOV-2001; 2001US-0331287P.				
XX	(HUMA-) HUMAN GENOME SCI INC.				
PA	Rosen CA, Ruben SM;				
XX	WPI; 2003-175238/17.				
XX	New human secreted proteins and nucleic acid molecules, useful for				
PT	preparing a diagnostic or pharmaceutical composition for diagnosing,				
PT	preventing or treating cancer or other hyperproliferative disorder,				
PT	asthma, allergies or AIDS.				
XX	Claim 9; SEQ ID NO 153; 3205pp; English.				
PS	The invention relates to novel genes ADA39629-ADA40565 and proteins				
XX	ADA40566-ADA41501 for human secreted proteins, useful for preventing,				
CC	treating or ameliorating medical conditions e.g. by protein or gene				
CC	therapy. The polypeptides, nucleic acid molecules, antibodies or their				

CC fragments, and agonists or antagonists that bind to the polypeptide are
CC useful for preparing a diagnostic or pharmaceutical composition for
CC diagnosing or treating cancer or other hyperproliferative disorder. The
CC polypeptides and nucleic acid molecules are also useful for detecting,
CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
CC or other hyperproliferative disorders including neoplasms, autoimmune
CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
CC thrombocytopenia), allergic reactions including asthma or eczema,
CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
CC Alzheimer's disease or Parkinson's disease) cardiovascular disorders
CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
CC fungal or viral infections including HIV/AIDS), or wound healing and
CC disorders of epithelial cell proliferation. The nucleic acids are also
CC useful for chromosome identification, radiation hybrid mapping or long-
CC range restriction mapping, as molecular weight markers, or as
CC hybridization or diagnostic probes. The polypeptides and antibodies are
CC useful for providing immunological probes for differential identification
CC of the tissues immunohistochemistry assays. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;

Query Match 100.0%; Score 1434; DB 7; Length 1434;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATTAAATCTTTTATCGGGAATGATGATATATTTCAATGTCACATTCATGTTGA 60
DB 1 CATTAAATCTTTTATCGGGAATGATGATATATTTCAATGTCACATTCATGTTGA 60
QY 61 TTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAAATTTTATATTTGAAAGTCAGTGC 120
DB 61 TTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAAATTTTATATTTGAAAGTCAGTGC 120
QY 121 AAAATTTAGATAGATATATAATAATAACAAGTAATAACAAGTCAAGCAGTGT 180
DB 121 AAAATTTAGATAGATATATAATAATAACAAGTAATAACAAGTCAAGCAGTGT 180
QY 181 CTAATAAATAATCTGGGTTCTTAAATAATTTTAAATTTATCTGAAATAGTTTCT 240
DB 181 CTAATAAATAATCTGGGTTCTTAAATAATTTTAAATTTATCTGAAATAGTTTCT 240
QY 241 TAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCAATAAA 300
DB 241 TAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCAATAAA 300
QY 301 CAAATTTGCTATTAATTCGACAGTGTGTAATATACAGAAATTTATCAGGCATTTACCAAGTC 360
DB 301 CAAATTTGCTATTAATTCGACAGTGTGTAATATACAGAAATTTATCAGGCATTTACCAAGTC 360
QY 361 TAGGCACATATAGGAATGACAGCACTCAGAAATGTTTCAATGTAGTATGCTGCTGTA 420
DB 361 TAGGCACATATAGGAATGACAGCACTCAGAAATGTTTCAATGTAGTATGCTGCTGTA 420
QY 421 AGGTAGGGGAGCTTATTACAGATAGTAGATAGTTTCTTAATGCTGTSTCAATTTGCTGG 480
DB 421 AGGTAGGGGAGCTTATTACAGATAGTAGATAGTTTCTTAATGCTGTSTCAATTTGCTGG 480
QY 481 CCTTTGGCTACTGTACTTCCSCATATGCGAGCCCATTCAGTCTTGAGTTTCTTCTCT 540
DB 481 CCTTTGGCTACTGTACTTCCSCATATGCGAGCCCATTCAGTCTTGAGTTTCTTCTCT 540
QY 541 GGACACCTTATGCTCTGAAATCATGAGGAGGCTGATTCAAATTTGGTGAATTTGGGTAGAAA 600
DB 541 GGACACCTTATGCTCTGAAATCATGAGGAGGCTGATTCAAATTTGGTGAATTTGGGTAGAAA 600
QY 601 GCAGTATGTTTGTGCAATTAAGATGATAGTTTATAGATAGTTTATAGCTTTAGTGTAT 660
DB 601 GCAGTATGTTTGTGCAATTAAGATGATAGTTTATAGATAGTTTATAGCTTTAGTGTAT 660

DB 601 GCAGTATGTTTGTGCTGACATTAAGATGCTAGGTATATAGATAGTTTATAGCTTTAAGTGTAT 660
QY 661 GTTTTATATCTTTAAATAAAGAAATATAACCTTTTAAAGCTATTCCACTCTCTCCCCGAGC 720
DB 661 GTTTTATATCTTTAAATAAAGAAATATAACCTTTTAAAGCTATTCCACTCTCTCCCCGAGC 720
QY 721 CTATCTCAAACTGCTGGAATATATGAGAGATCTTTGAAAGAAAGTAAATAATAACCTTCACT 780
DB 721 CTATCTCAAACTGCTGGAATATATGAGAGATCTTTGAAAGAAAGTAAATAATAACCTTCACT 780
QY 781 GCTCCACTCCAGGTGAATCCGCCCACTCCCACTGACCTAGTAGAATTTGTAATTTAATAC 840
DB 781 GCTCCACTCCAGGTGAATCCGCCCACTCCCACTGACCTAGTAGAATTTGTAATTTAATAC 840
QY 841 TTACCTTCTATTTCTGAAATCAGTTGTGAACCTTGTGCTTATGTTTCAGARGTTTAAAGAC 900
DB 841 TTACCTTCTATTTCTGAAATCAGTTGTGAACCTTGTGCTTATGTTTCAGARGTTTAAAGAC 900
QY 901 CTCMGTCGAATTCATTTTAAATCTGCTATTCTGAGAAGCATTGAATCTTAAACA 960
DB 901 CTCMGTCGAATTCATTTTAAATCTGCTATTCTGAGAAGCATTGAATCTTAAACA 960
QY 961 AGAAGACTCATCTGCTAGCTGTTGCTCACTCTATGAGCCCAATAAGGGTTCTGTGCTTA 1020
DB 961 AGAAGACTCATCTGCTAGCTGTTGCTCACTCTATGAGCCCAATAAGGGTTCTGTGCTTA 1020
QY 1021 GCATTAACAAATAAGGTTTATAGTTAAAGCCCAATGATTAATTTTTCGATGGAGG 1080
DB 1021 GCATTAACAAATAAGGTTTATAGTTAAAGCCCAATGATTAATTTTTCGATGGAGG 1080
QY 1081 GCTTTAAATTTGTGCTCTTTTTCATATTTTATTCATATTCATTTATGTTTCTAACTG 1140
DB 1081 GCTTTAAATTTGTGCTCTTTTTCATATTTTATTCATATTCATTTATGTTTCTAACTG 1140
QY 1141 CTTTTAGGAGATATAATATATGTTTAAATTTAGTTTGGGGGAAATAATTTGTCGAAAGA 1200
DB 1141 CTTTTAGGAGATATAATATATGTTTAAATTTAGTTTGGGGGAAATAATTTGTCGAAAGA 1200
QY 1201 GGATAATTAATTTACCTGCTCTGTTTATTCAGATAAAGAGAGAGAGTACGCTGCATA 1260
DB 1201 GGATAATTAATTTACCTGCTCTGTTTATTCAGATAAAGAGAGAGAGTACGCTGCATA 1260
QY 1261 TTCAAGAGTTGTACCTTAAACATTCGTTGAAACATTTTCTAAGATTTTCAAAAGGAATAT 1320
DB 1261 TTCAAGAGTTGTACCTTAAACATTCGTTGAAACATTTTCTAAGATTTTCAAAAGGAATAT 1320
QY 1321 GTGTAAATTTGAGAAATCATACCACTGCTTAAACCACTGCTTAAACCACTGCTTAAATTA 1380
DB 1321 GTGTAAATTTGAGAAATCATACCACTGCTTAAACCACTGCTTAAACCACTGCTTAAATTA 1380
QY 1381 AGTATTTAAATGATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 1434
DB 1381 AGTATTTAAATGATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTT 1434

RESULT 2
ACC50424

ID ACC50424 standard; cDNA; 1434 BP.

XX ACC50424;

XX 12-JUN-2003 (first entry)

XX Human secreted protein coding sequence, SEQ ID 91.

XX Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
XX vulnery; antiinflammatory; nootropic; neuroprotective;
KW antiparkinsonian; gene therapy; human; cardiovascular disorder; gene; ss.

XX Homo sapiens.

XX WO200295010-A2.

XX

PD 28-NOV-2002.

XX 19-MAR-2002; 2002WO-US009785.

XX 21-MAR-2001; 2001US-0277340P.

XX 19-JUL-2001; 2001US-0306171P.

XX 13-NOV-2001; 2001US-0331287P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI; 2003-129429/12.

XX Novel human secreted proteins, useful for detecting, preventing,

PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular

PT disorders such as arrhythmia.

XX Claim 21; SEQ ID NO 91; 1881pp; English.

XX The present invention relates to novel human secreted proteins (ABR47633-
ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins
CC and their coding sequences are useful for the preparation of a diagnostic
CC or pharmaceutical composition for diagnosing or treating a cardiovascular
CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary
CC arteriosclerosis and myocardial ischaemia), neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, renal disorders,
CC proliferative disorders and/or cancerous diseases and conditions, for
CC wound healing and epithelial cell proliferation, to treat inflammation or
CC infection, for treating thrombosis and arteriosclerosis, for treating or
CC preventing neural damage which occurs in neuronal disorders or
CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's
CC disease, to enhance bone and periodontal regeneration and aid in tissue
CC transplants or bone grafts, to prevent skin aging or hair loss, to
CC stimulate growth and differentiation of haematopoietic cells and bone
CC marrow cells when used in combination with other cytokines, to maintain
CC organs before transplantation or for supporting cell culture of primary
CC tissues, to increase or decrease differentiation or proliferation of
CC embryonic stem cells, or to modulate mammalian characteristics or
CC metabolism. Note: The sequence data for this patent was published in
CC electronic format and is available from WIPO at
CC ftp.wipo.int/pub/published_pot_sequences

XX Sequence 1434 BP; 480 A; 203 C; 250 G; 496 T; 0 U; 5 Other;

Query Match 100.0%; Score 1434; DB 7; Length 1434;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CATTAAGCTCTTTTATCGGGAATAGTATGATATTTTCAATGACATCCCATTCATGTTGA 60

DB 1 CATTAAGCTCTTTTATCGGGAATAGTATGATATTTTCAATGACATCCCATTCATGTTGA 60

QY 61 TTGGAGCTCACAGTATTTTGTGTAAGCAGAGATTTTATATTTTATATTTGAAATCAGTGC 120

DB 61 TTGGAGCTCACAGTATTTTGTGTAAGCAGAGATTTTATATTTTATATTTGAAATCAGTGC 120

QY 121 AAAAATTATGAATAGGATATATCTAATAATACAAAGTAAATACAAAGTCAAGCAGTGT 180

DB 121 AAAAATTATGAATAGGATATATCTAATAATACAAAGTAAATACAAAGTCAAGCAGTGT 180

QY 181 CTAATAAAAAATTCGGTTCCTTTAAAAATTTTAAAAATTTTATCTTGAATAGTTTCT 240

DB 181 CTAATAAAAAATTCGGTTCCTTTAAAAATTTTAAAAATTTTATCTTGAATAGTTTCT 240

QY 241 TAGAATAATCTCAGGATATCAGAAAGTCAATTAAGTGTGATAGTATATCAATATAA 300

DB 241 TAGAATAATCTCAGGATATCAGAAAGTCAATTAAGTGTGATAGTATATCAATATAA 300

QY 301 CAAATGCTCTATTAATGCAAGAGTGGTAAATATACAGAAATTTATCAGGCATTTACCAAGTC 360

DB 301 CAAATGCTCTATTAATGCAAGAGTGGTAAATATACAGAAATTTATCAGGCATTTACCAAGTC 360

QY 361 TAGGCACATATAGGAAATGAGCACTCAGAAATGGTTTCAATGTAGTAGTGTGATGCTTGTGA 420

DB 361 TAGGCACATATAGGAAATGAGCACTCAGAAATGGTTTCAATGTAGTAGTGTGATGCTTGTGA 420

QY 421 AGGTAGGGGAGCTTTATTCAGACATAGTAGTATGTTTCTTAATGCTGTSTCAATTTGCTGG 480

DB 421 AGGTAGGGGAGCTTTATTCAGACATAGTAGTATGTTTCTTAATGCTGTSTCAATTTGCTGG 480

QY 481 CCTTTGGCTACCTGTACTTCCSCAATATGCGAGCCCATTCAGTCTTGAGTGTCTTCTCTCT 540

DB 481 CCTTTGGCTACCTGTACTTCCSCAATATGCGAGCCCATTCAGTCTTGAGTGTCTTCTCTCT 540

QY 541 GGACACCTTATGCTCTGAAATCATGAGCGAGGCTGATTCAATTTGGTGTAGGAGAAA 600

DB 541 GGACACCTTATGCTCTGAAATCATGAGCGAGGCTGATTCAATTTGGTGTAGGAGAAA 600

QY 601 GCAGTATGTTTGTGCTGACATTAAGATGTAGGTATAGATAGGTTTACGCTTTAAAGTGTAT 660

DB 601 GCAGTATGTTTGTGCTGACATTAAGATGTAGGTATAGATAGGTTTACGCTTTAAAGTGTAT 660

QY 661 GTTTTATACCTTTAAATTAAGAAATATAACCTTTTAAGCTATTCACCTCTCTCCCGCAGC 720

DB 661 GTTTTATACCTTTAAATTAAGAAATATAACCTTTTAAGCTATTCACCTCTCTCCCGCAGC 720

QY 721 CTATCTCAAACTGCTGGAATATATGAGAGAGATCTTGAAAGAAATATAAATTAACCTTCACT 780

DB 721 CTATCTCAAACTGCTGGAATATATGAGAGAGATCTTGAAAGAAATATAAATTAACCTTCACT 780

QY 781 GCTCCACTCCAGGTGAATCCGCCACTCCCACTCCCACTAGTAGTAATTTGTAATTTAATAC 840

DB 781 GCTCCACTCCAGGTGAATCCGCCACTCCCACTCCCACTAGTAGTAATTTGTAATTTAATAC 840

QY 841 TTACCTCTATTTCTGAAATCAGTTGGAATGTTGCTTATGTTTCAAGAGTTTAAAGAAC 900

DB 841 TTACCTCTATTTCTGAAATCAGTTGGAATGTTGCTTATGTTTCAAGAGTTTAAAGAAC 900

QY 901 CTCMGTAATTCATTTTTTAAAAATCTGCTATTCGAGAAGCATTTGAATTTGTAATTTAACA 960

DB 901 CTCMGTAATTCATTTTTTAAAAATCTGCTATTCGAGAAGCATTTGAATTTGTAATTTAACA 960

QY 961 AGAAGACTCATCTGTAGTCTTTGCTGACTCTCTATGAGCCCCCATTAAGGGTTCTGTGCTTA 1020

DB 961 AGAAGACTCATCTGTAGTCTTTGCTGACTCTCTATGAGCCCCCATTAAGGGTTCTGTGCTTA 1020

QY 1021 GCATTAACAAATAAGGTTTATAGGTAAAGCCCAATGTTATTAATTTTGTGATGAGG 1080

DB 1021 GCATTAACAAATAAGGTTTATAGGTAAAGCCCAATGTTATTAATTTTGTGATGAGG 1080

QY 1081 GCTTTAAAAATTTGCTCTCTTTTCAATTTTATTCATATTTCAATTTATGTTTGTAACTG 1140

DB 1081 GCTTTAAAAATTTGCTCTCTTTTCAATTTTATTCATATTTCAATTTATGTTTGTAACTG 1140

QY 1141 CTTTTAGGGAGATAATATATGTTATAATTAAGTTTGGGGGGAATAATTTGTGCAAGA 1200

DB 1141 CTTTTAGGGAGATAATATATGTTATAATTAAGTTTGGGGGGAATAATTTGTGCAAGA 1200

QY 1201 GGATAAATTAATTTAGCTGCTCTGTTTATTCAGAATAAAGAGAGAGACTACCGTCATA 1260

DB 1201 GGATAAATTAATTTAGCTGCTCTGTTTATTCAGAATAAAGAGAGAGACTACCGTCATA 1260

QY 1261 TTCAGAGTTGTACCTTTAAACATTTGSGTGAACATTTTTTCTAAGATTTTCAAAAAGGAATAT 1320

DB 1261 TTCAGAGTTGTACCTTTAAACATTTGSGTGAACATTTTTTCTAAGATTTTCAAAAAGGAATAT 1320

QY 1321 GTGTAATTTGAGAAATCATAAACACCTGTCTTAATCTTGTTAAACAAACTGTCTTCTTAATAA 1380

DB 1321 GTGTAATTTGAGAAATCATAAACACCTGTCTTAATCTTGTTAAACAAACTGTCTTCTTAATAA 1380

QY 1381 AGTATTTAATGATTTTTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1434

DB 1381 AGTATTTAATGATTTTTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1434

RESULT 3
AAX37452

ID AAX37452 standard; cDNA; 1434 BP.

XX AAX37452;
XX 06-JUL-1999 (first entry)

DT Human secreted protein cDNA fragment containing gene 2.
XX

DE Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection; ss.

XX Homo sapiens.
XX

XX WO9918208-A1.
XX

XX 15-APR-1999.
XX

XX 01-OCT-1998; 98WO-US020775.
XX

XX 02-OCT-1997; 97US-0060833P.
XX

XX 02-OCT-1997; 97US-0060836P.
XX

XX 02-OCT-1997; 97US-0060837P.
XX

XX 02-OCT-1997; 97US-0060838P.
XX

XX 02-OCT-1997; 97US-0060839P.
XX

XX 02-OCT-1997; 97US-0060843P.
XX

XX 02-OCT-1997; 97US-0060862P.
XX

XX 02-OCT-1997; 97US-0060866P.
XX

XX 02-OCT-1997; 97US-0060874P.
XX

XX 02-OCT-1997; 97US-0060880P.
XX

XX 02-OCT-1997; 97US-0060884P.
XX

XX (HUMA-) HUMAN GENOME SCI INC.
XX

XX Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;
PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;
PI Lafleur DW, Shi Y;

XX WPI; 1999-264022/22.
XX

XX P-PSDB; AAY07853.
XX

XX New isolated human genes and the secreted polypeptides they encode.
XX

XX Claim 1a; Page 228; 368pp; English.
XX

XX This invention describes novel isolated human genes and the secreted
CC proteins they encode. The products of the invention are useful for
CC preventing, treating or ameliorating medical conditions, e.g. by protein
CC or gene therapy. Also pathological conditions can be diagnosed by
CC determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 101 polynucleotides, based on
CC which tissues they are most highly expressed in, and include developing
CC products for the diagnosis or treatment of cancer, tumours,
CC neurodegenerative disorders, developmental abnormalities and fetal
CC deficiencies, blood disorders, leukemias, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
CC transplant rejection, disorders involving osteoclasts such as
CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
CC infections and AIDS. The human secreted proteins of the invention are
CC represented in AAY07852-Y07993 and the encoding nucleic acids are
CC represented in AAX37451-X37552

SQ	Sequence	1434 BP; 480 A; 204 C; 250 G; 495 T; 0 U; 5 Other;
	Query Match	96.4%; Score 1383; DB 2; Length 1434;
	Best Local Similarity	99.9%; Pred. No. 0;
	Matches 1433; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1	CAATTAACCTCTTTTATCGGGAATAGTATGATATTTCAATGTGCACCTCCATTCAATGTTGA 60
DB	1	CAATTAACCTCTTTTATCGGGAATAGTATGATATTTCAATGTGCACCTCCATTCAATGTTGA 60
QY	61	TTTGGAGCTGACAGTATTTTGTGTAAGCAGAGATTTAAATTTTATATTTGAAAGTCAGTGC 120
DB	61	TTTGGAGCTGACAGTATTTTGTGTAAGCAGAGATTTAAATTTTATATTTGAAAGTCAGTGC 120
QY	121	AAATTTATGATAGGATATCTAATAATCAAGTAAATCAAAAGTCAAAAGTCAAAAGTCAAGTGT 180
DB	121	AAATTTATGATAGGATATCTAATAATCAAGTAAATCAAAAGTCAAAAGTCAAAAGTCAAGTGT 180
QY	181	CTAATAAAAAATCTGGGTTCTTAAAAATTTATTTAAATTTATCTTGAATAGTCTTCT 240
DB	181	CTAATAAAAAATCTGGGTTCTTAAAAATTTATTTAAATTTATCTTGAATAGTCTTCT 240
QY	241	TAGATTAATCTCAGGATATGAGAAATCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 300
DB	241	TAGATTAATCTCAGGATATGAGAAATCAATTAAGTGTGAGTAAAGTTAGTATCATTTAA 300
QY	301	CAAAATTTCTAATAATGCAMGAGTGTATATATACAGAAATTTATCAGGCATTTACCAAGTC 360
DB	301	CAAAATTTCTAATAATGCAMGAGTGTATATATACAGAAATTTATCAGGCATTTACCAAGTC 360
QY	361	TAGGCACATATAGGAAATGCAGCACTCAGAAATGCTTCAATGTAGTAGTTGATGCTTGA 420
DB	361	TAGGCACATATAGGAAATGCAGCACTCAGAAATGCTTCAATGTAGTAGTTGATGCTTGA 420
QY	421	AGGTAGGGAGCTTATTCAGACATAGTAGATAGTTTCTTAATGCTGTSTCAATTCGTGG 480
DB	421	AGGTAGGGAGCTTATTCAGACATAGTAGATAGTTTCTTAATGCTGTSTCAATTCGTGG 480
QY	481	CTTTTGCTACCTGCTACTTCCSCATTTATGGCAGCCCACTCAGTCTTGAGTTTCTTCTCT 540
DB	481	CTTTTGCTACCTGCTACTTCCSCATTTATGGCAGCCCACTCAGTCTTGAGTTTCTTCTCT 540
QY	541	GGACACCTTATGCTCTGAAATCATGAGCAGGCTGATTCATTTGGTGAATTTGGGTAGAGA 600
DB	541	GGACACCTTATGCTCTGAAATCATGAGCAGGCTGATTCATTTGGTGAATTTGGGTAGAGA 600
QY	601	GCAGTATGTTTGTCTGACATTAAGATGTAGTATAGTATAGTGTAGCTTTAAGTGTAT 660
DB	601	GCAGTATGTTTGTCTGACATTAAGATGTAGTATAGTGTAGCTTTAAGTGTAT 660
QY	661	GTTTTATACCTTTAAATAAGAAATATAACCTTTTAAAGTATTCACCTCTCCCTCCACG 720
DB	661	GTTTTATACCTTTAAATAAGAAATATAACCTTTTAAAGTATTCACCTCTCCCTCCACG 720
QY	721	CTATCTCAAACTGGTGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCACT 780
DB	721	CTATCTCAAACTGGTGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCACT 780
QY	781	GCTCAGCTCAGGAGTAAATCCGCCCACTCCCACTGACCTAGTAGTAATTTGTAATTAATAC 840
DB	781	GCTCAGCTCAGGAGTAAATCCGCCCACTCCCACTGACCTAGTAGTAATTTGTAATTAATAC 840
QY	841	TTACTCTTATTTCTGAAATCAGTTGTGAACCTGTTGCTTATGTTTCAAGTGTAAAGAAC 900
DB	841	TTACTCTTATTTCTGAAATCAGTTGTGAACCTGTTGCTTATGTTTCAAGTGTAAAGAAC 900
QY	901	CTCMGTGAATTCATTTTAAAAATCTGCTATCTGAGAAAGCAATTTGAATTAATCTTAACA 960
DB	901	CTCMGTGAATTCATTTTAAAAATCTGCTATCTGAGAAAGCAATTTGAATTAATCTTAACA 960
QY	961	AGAAGACTCATCTGTAGTGTGTTCTGACTCTCCTATGAGCCCAATAGGGGTCTGTCGCTTA 1020
DB	961	AGAAGACTCATCTGTAGTGTGTTCTGACTCTCCTATGAGCCCAATAGGGGTCTGTCGCTTA 1020

PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
 XX WPI; 2000-293155/25.
 XX Polynucleotide library comprising 1079 defined sequences, useful in the
 PT form of an array to detect cancer or susceptibility to cancer.
 XX
 XX Claim 1; Page 268; 502pp; English.
 XX
 XX The present invention describes a library of polynucleotides comprising
 CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
 CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
 CC one of the 1079 sequences; (2) a recombinant host cell containing (I);
 CC (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
 CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
 CC of detecting differentially expressed genes correlated with a cancerous
 CC state of a mammalian cell comprising detecting a gene product encoded by
 CC 65 of the 1079 sequences given in the specification. The polynucleotides
 CC are used to monitor patients having (or susceptible) to cancer to detect
 CC potentially malignant events at a molecular level before they are
 CC detectable at a gross morphological level. The polynucleotides are also
 CC useful for monitoring the efficacy of various therapies and preventive
 CC interventions. Polynucleotide probes based on the disclosed sequences are
 CC useful for chromosome mapping and detection of transcription levels. The
 CC 1079 polynucleotide sequences were derived from a human colon cancer cell
 CC line Kml2L4-A cDNA library
 XX
 XX Sequence 415 BP; 114 A; 73 C; 87 G; 141 T; 0 U; 0 Other;
 SQ

Query Match 10.1%; Score 145; DB 3; Length 415;
 Best Local Similarity 100.0%; Pred. No. 4.7e-41;
 Matches 145; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 324 GTGGTAATATACAGAAATTTATCAGGCATTACCAAGCTCTAGGCACATATAGGAAATGCAGC 383
 Db 24 GTGGTAATATACAGAAATTTATCAGGCATTACCAAGCTCTAGGCACATATAGGAAATGCAGC 83

Qy 384 ACTCAGATGGTTTCAAGTAGTAGTTGATGCTTGTAGGTAGGGAGCTTATTCAGACA 443
 Db 84 ACTCAGATGGTTTCAAGTAGTAGTTGATGCTTGTAGGTAGGGAGCTTATTCAGACA 143

Qy 444 TAGTAGATAGTTTCTCTAATGCTGT 468
 Db 144 TAGTAGATAGTTTCTCTAATGCTGT 168

RESULT 6
 AAC10084
 ID AAC10084 standard; cDNA; 395 BP.
 AC AAC10084;
 XX
 XX 06-OCT-2000 (first entry)
 DT
 XX Human secreted protein 5' EST, SEQ ID NO: 14159.
 DE
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.
 KW
 XX Homo sapiens.
 OS
 XX EP1033401-A2.
 FN
 XX 06-SEP-2000.
 PD
 XX 21-FEB-2000; 2000EP-00200610.
 PF
 XX 26-FEB-1999; 99US-0122487P.
 PR
 XX (GEST) GENSET.
 XX
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 PI WPI; 1995-206931/27.
 XX

DR WPI; 2000-500381/45.
 XX
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX
 XX Claim 1; SEQ ID NO 14159; 71pp + Sequence Listing; English.
 XX
 XX The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
 CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used
 CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors
 XX
 XX Sequence 395 BP; 147 A; 41 C; 74 G; 125 T; 0 U; 8 Other;
 SQ

Query Match 9.9%; Score 142; DB 3; Length 395;
 Best Local Similarity 100.0%; Pred. No. 5.2e-40;
 Matches 142; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 TGATTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAATTTTATTTGAAAGTCAG 117
 Db 1 TGATTTGGAGCTGACAGTTATTTTGTGTAAGCAGAGATTTAATTTTATTTGAAAGTCAG 60

Qy 118 TGC AAAATTATGATAGATATATACTAATAATACAAAGTAAACAAAGTCAAAGCAGT 177
 Db 61 TGC AAAATTATGATAGATATATACTAATAATACAAAGTAAACAAAGTCAAAGCAGT 120

Qy 178 GTTCTAAATAAAAATTCCTGGGT 199
 Db 121 GTTCTAAATAAAAATTCCTGGGT 142

RESULT 7
 AAT25136
 ID AAT25136 standard; DNA; 336 BP.
 XX
 XX AAT25136;
 AC
 XX 22-OCT-1996 (first entry)
 DT
 XX Human gene signature HUMGS07294.
 DE
 XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 KW cell typing; abnormal cell function; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO9514772-A1.
 FN
 XX 01-JUN-1995.
 PD
 XX 11-NOV-1994; 94WO-JP001916.
 PF
 XX 12-NOV-1993; 93JP-00355504.
 PR
 XX (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 XX
 XX Matsubara K, Okubo K;
 PI WPI; 1995-206931/27.
 XX

PT Single-stranded DNA for identifying gene signatures - isolated from 3'-
 PT directed human cDNA library that reflects relative abundance of corresp.
 XX mRNA in specific human tissues.

PS Claim 1; Page 1777-78; 2245pp; Japanese.

XX A single-stranded DNA (or its complementary strand or the corresp. double
 CC -stranded DNA) which comprises one of the 7837 "GS" sequences given in
 CC AAT19001-726937 and which is able to hybridise to part of human genomic
 CC DNA, cDNA or mRNA is claimed. The GS (Gene Signature) sequences were
 CC obtained from 3'-directed cDNA libraries prepared from various human
 CC tissues; synthesis of cDNA was initiated from the 3'-end of mRNA by using
 CC poly(T) as the sole primer. Since the 3'-untranslated sequence is unique
 CC to a particular mRNA species, almost all the 3'-oriented cDNAs hybridise
 CC with specific mRNAs. Each library is constructed so as to reflect
 CC accurately the relative abundance of different mRNAs in the particular
 CC tissue from which it was derived. The appearance frequency of a given GS
 CC in a cDNA library can be determined (esp. using primers and probes
 CC derived from the GS sequences) as a means of diagnosing abnormal cell
 CC function or for recognising different cell types

XX Sequence 336 BP; 95 A; 62 C; 62 G; 112 T; 0 U; 5 Other;

Query Match 8.9%; Score 127; DB 2; Length 336;

Best Local Similarity 100.0%; Pred. No. 8.7e-35;

Matches 127; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 763 GTAAATAAACCTTCACCTGCTCCACTCCAGTGTAATCGGCCACTCCCACTGACCTAGTA 822

Db 14 GTAAATAAACCTTCACCTGCTCCACTCCAGTGTAATCGGCCACTCCCACTGACCTAGTA 73

QY 823 GAATTCGTAATTTAATACCTTCTTCTGAAATCAGTTGTAACCTGTCCTTAT 882

Db 74 GAATTCGTAATTTAATACCTTCTTCTGAAATCAGTTGTAACCTGTCCTTAT 133

QY 883 GTTCAGA 889

Db 134 GTTCAGA 140

RESULT 8

AD881840/c

ID ADE81840 standard; cDNA; 493 BP.

XX ADE81840;

XX 29-JAN-2004 (first entry)

XX Arabidopsis thaliana expressed polynucleotide seq id 611.

XX genetically modified organism; transgenic organism; plant;
 KW inhibitor testing; activator testing; modifier testing; fungicide;
 KW insecticide; genetic function; genetic regulation; cellular metabolism;
 KW gene; ss.

XX Arabidopsis thaliana.

XX US2003115639-A1.

XX 19-JUN-2003.

XX 26-JAN-2001; 2001US-00770961.

XX 27-JAN-2000; 2000US-0178466P.

XX (GORL/) GORLACH J.

XX (ANY/) AN Y.

XX (HAMI/) HAMILTON C M.

XX (PRIC/) PRICE J L.

XX (RAIN/) RAINES T M.

XX (YUY/) YU Y.

XX (RAME/) RAMEKA J G.

XX (PAGE/) PAGE A.

PA (MATH/) MATHEW A V.
 PA (LEDF/) LEDFORD B L.
 PA (WOES/) WOESSNER J P.

PA (HAAS/) HAAS W D.

PA (GARC/) GARCIA C A.

PA (KRIC/) KRICKER M.

PA (SLAT/) SLATER T.

PA (DAVI/) DAVIS K R.

PA (ALLE/) ALLEN K.

PA (HOFF/) HOFFMAN N.

PA (HURB/) HURBAN P.

XX

PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;

PI Rameka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;

PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;

PI Hurban P;

XX

DR WPI; 2003-810930/76.

XX Novel Arabidopsis thaliana nucleic acids useful for generating

PT genetically modified transgenic organisms, for screening biologically

PT active agents such as fungicides, insecticides.

XX Claim 1; SEQ ID NO 611; 44pp; English.

PS The invention describes a nucleic acid (I) comprising a sequence capable

XX of hybridising under stringent conditions to any one of 999 fully defined

CC Arabidopsis thaliana sequences (I) as given in specification e.g., 360,

CC 1137, 455, 219, 472 nucleotides, etc, or its fragment. (I) is useful as a

CC hybridisation probe to complementary molecules in a cDNA library. (I) is

CC also useful for generating genetically modified and transgenic organisms,

CC usually plant cells and plants. A protein encoded by (I) is useful in

CC screening assays to determine the effect of candidate inhibitors,

CC activators or modifiers of the gene product. The protein is also useful

CC for screening biologically active agents e.g., fungicides and

CC insecticides. A genetically modified cell, comprising an exogenous

CC nucleic acid, where the nucleic acid comprises transcription regulatory

CC sequences operably linked to a sequence capable of hybridising under

CC stringent conditions to (I) is useful in the study of genetic function

CC and regulation, for alteration of the cellular metabolism and for

CC screening compounds that may affect the biological function of the gene

CC or gene product. This sequence represents an Arabidopsis thaliana

CC polynucleotide of the invention.

XX

SQ Sequence 493 BP; 146 A; 115 C; 70 G; 162 T; 0 U; 0 Other;

Query Match 3.1%; Score 45; DB 9; Length 493;

Best Local Similarity 100.0%; Pred. No. 2.4e-06;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA 1434

Db 52 TGATTTTAAAAA 8

RESULT 9

AAH25190

ID AAH25190 standard; cDNA; 1231 BP.

XX

AC AAH25190;

XX

DT 22-AUG-2001 (first entry)

XX

DE Nucleotide sequence of a human transferrin.

XX

KW Human; transferrin; autoimmune disease; rheumatoid arthritis;

KW hyperproliferative disorder; neoplasm; cardiovascular disorder;

KW cardiac arrest; cerebrovascular disorder; cerebral ischemia;

KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;

KW ocular disorder; corneal infection; wound healing;

KW epithelial cell proliferation; aging; organ transplant; ss.

XX Homo sapiens.

OS

```
XX Key Location/Qualifiers
FH 3.989
FT /*tag= a
FT /product= "transferrin"
XX WO200146254-A1.
XX 28-JUN-2001.
XX 21-DEC-2000; 2000WO-US034769.
XX 23-DEC-1999; 99US-0171595P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Shi Y, Choi GH;
XX WPI; 2001-381910/40.
XX P-PSDB; AAB84371.
XX Isolated nucleic acid molecule encoding a human transferrin protein is
XX used in preventing, treating or ameliorating a medical condition.
XX Claim 1; Page 285; 290pp; English.
XX The present sequence encodes a human transferrin polypeptide. Transferrin
XX polypeptides and polynucleotides are used to prevent, treat or ameliorate
XX a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
XX dogs, chickens or sheep. Disorders which are diagnosed or treated include
XX autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,
XX infections caused by bacteria, viruses and fungi and ocular disorders
XX e.g. corneal infection. The polypeptides can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities
XX SQ Sequence 1231 BP; 408 A; 254 C; 305 G; 264 T; 0 U; 0 Other;
Query Match 3.1%; Score 45; DB 4; Length 1231;
Best Local Similarity 100.0%; Pred. No. 2e-06;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1390 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
DB 1147 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1191
RESULT 10
AAZ24411
ID AAZ24411 standard; cDNA; 1348 BP.
XX AAZ24411;
XX 14-FEB-2000 (first entry)
XX Human bladder tumour cDNA library derived EST 23.
XX Expressed sequence tag; human; bladder; tumour; cancer; cytostatic;
XX treatment; gene therapy; EST; ss.
XX Homo sapiens.
XX DE19818619-A1.
XX 28-OCT-1999.
XX 21-APR-1998; 98DE-01018619.
```

```
XX 21-APR-1998; 98DE-01018619.
XX (META-) METAGEN GES GENOMPORSCHUNG MBH.
XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX WPI; 1999-612028/53.
XX New nucleic acid sequences expressed in bladder tumor tissue, and derived
XX polypeptides, for treatment of bladder tumor and identification of
XX therapeutic agents.
XX Claim 3; Page 78; 132pp; German.
XX This invention describes novel polypeptide fragments (I) and the
XX polynucleotides (II) that encode them that are highly expressed in a
XX human bladder tumour and which have cytostatic activity. (II) are used
XX for recombinant expression of (I) and to isolate complete genes. (I) are
XX used to identify agents suitable for treatment of bladder cancer, to
XX directly treat this form of cancer (including expression from gene
XX therapy vectors) or are used in a preparation for cancer treatment. (I)
XX is also used for the generation of specific antibodies. (II) are
XX identified by assembling ESTs (expressed sequence tags) from a particular
XX tissue type before comparison of expression patterns. This allows a
XX significantly longer fragment of the gene to be revealed, and therefore
XX reduces the number of failures associated with the fact that ESTs from
XX different libraries may represent different parts of the same unknown
XX gene, distorting the estimated frequency of occurrence in a particular
XX tissue. AAZ3260-Z43309 represent expressed sequence tag (EST) fragments
XX isolated from a human bladder tumour cDNA library which encode the
XX proteins represented in AAY66143-Y66198
XX SQ Sequence 1348 BP; 326 A; 386 C; 349 G; 287 T; 0 U; 0 Other;
Query Match 3.1%; Score 45; DB 2; Length 1348;
Best Local Similarity 100.0%; Pred. No. 1.9e-06;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1390 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
DB 1299 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1343
RESULT 11
ABL89915
ID ABL89915 standard; cDNA; 2552 BP.
XX ABL89915;
XX 24-MAY-2002 (first entry)
XX Human polynucleotide SEQ ID NO 477.
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX neurological disease; infection; human; secreted protein; gene; ss.
XX OS Homo sapiens.
XX WO200190304-A2.
XX 29-NOV-2001.
XX 18-MAY-2001; 2001WO-US016450.
XX 19-MAY-2000; 2000US-0205515P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Birse CE, Rosen CA;
```

XX WPI; 2002-122018/16.
 DR P-PSDB; ABB89506.
 XX
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders.
 XX
 XX Claim 4; SEQ ID NO 477; 2081pp + Sequence Listing; English.
 PS
 CC The invention relates to novel genes (ABL89449-ABL90853) and proteins
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2552 BP; 807 A; 438 C; 512 G; 794 T; 0 U; 1 Other;

Query Match 3.1%; Score 45; DB 6; Length 2552;
 Best Local Similarity 100.0%; Pred. No. 1.7e-06;
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1390 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
 DB 2504 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2548

RESULT 12
 ABV25066.
 ID ABV25066 standard; cDNA; 3470 BP.
 AC ABV25066;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 25057.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-0183131P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (WILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JE;
 XX
 DR WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 XX Claim 1; Page 4868-4869; 11750pp; English.
 PS
 CC The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 3470 BP; 1031 A; 694 C; 747 G; 938 T; 0 U; 60 Other;

Query Match 3.1%; Score 45; DB 5; Length 3470;
 Best Local Similarity 100.0%; Pred. No. 1.6e-06;
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1390 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
 DB 3313 TGATTTTAAAAA...AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3357

RESULT 13
 AAA87997
 ID AAA87997 standard; cDNA; 1730 BP.

AC AAA87997;
 XX
 DT 08-DEC-2000 (first entry)
 XX
 DE Torenia hybrida flavone synthase encoding cDNA SEQ ID NO:3.
 XX
 KW Flavone; flavanone; flavone synthase; plant; flower colour;
 KW bacterial resistance; nitrogen fixing bacteria; protection;
 KW ultraviolet radiation; ss.
 XX
 OS Torenia hybrida.

Key Location/Qualifiers
 CDS 57..1595
 /tag= a
 /product= "flavone synthase"

WO200044907-A1.
 PD 03-AUG-2000.
 XX
 PF 28-JAN-2000; 2000WO-JP000490.
 XX
 PR 29-JAN-1999; 99JP-00022427.
 PR 19-JUL-1999; 99JP-00205229.
 XX
 PA (SUNR) SUNTORY LTD.
 XX
 PI Mizutani M, Tanaka Y, Kusumi T, Ayabe S, Akashi T;
 XX
 XX WPI; 2000-543394/49.
 DR P-PSDB; AAB20573.
 XX
 DE Gene encoding a flavone synthase for production of Antirrhinum and other
 PT plants with modified flower colour.
 PT
 PS Claim 5; Page 34-38; 54pp; Japanese.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 21:11:14 , Search time 128 Seconds
(without alignments)
6217.184 Million cell updates/sec

Title: US-10-799-747-12
Perfect score: 1434
Sequence: 1 cattaaactcttttttcg.....aaaaaaaaaaaaaaaaaaaaa 1434

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/2/ina/5A COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B COMB.seq.*
5: /cgn2_6/prodata/2/ina/PCTUS COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfilee1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	3.1	1730	4	US-09-672-785-3
2	43	3.0	222	3	US-08-481-190-15
3	43	3.0	222	5	PCT-US93-00869-15
4	43	3.0	399	1	US-07-885-970A-13
5	43	3.0	399	1	US-08-298-687A-13
6	43	3.0	399	1	US-08-530-797-12
7	43	3.0	399	2	US-08-298-829-13
8	43	3.0	399	2	US-08-787-335-12
9	43	3.0	1114	4	US-09-152-060-41
10	43	3.0	1307	4	US-09-250-609-3
11	43	3.0	1307	4	US-09-250-611-3
12	43	3.0	1919	4	US-09-614-912-175
13	43	3.0	1976	4	US-09-920-759-10
14	43	3.0	2202	4	US-09-388-743-1
15	43	3.0	6200	4	US-09-439-923-1
16	43	3.0	6200	4	US-09-711-202A-1
17	43	3.0	6200	4	US-09-711-205A-1
18	42	2.9	144	1	US-08-702-344-26
19	42	2.9	258	4	US-09-621-976-15353
20	42	2.9	261	4	US-09-621-976-18330
21	42	2.9	276	4	US-09-621-976-18329
22	42	2.9	375	3	US-08-946-026-23
23	42	2.9	388	4	US-09-621-976-18573
24	42	2.9	396	4	US-09-640-173-10
25	42	2.9	396	4	US-09-640-173-42
26	42	2.9	396	4	US-09-713-550-10
27	42	2.9	396	4	US-09-713-550-42

```

28 42 2.9 769 4 US-08-567-882-5
29 42 2.9 911 2 US-08-924-759-9
30 42 2.9 911 3 US-08-248-335-9
31 42 2.9 991 3 US-08-924-747-25
32 42 2.9 991 3 US-09-247-373B-25
33 42 2.9 991 3 US-09-296-715-25
34 42 2.9 1035 2 US-08-580-545B-3
35 42 2.9 1035 3 US-09-262-653A-3
36 42 2.9 1035 3 US-08-867-484A-1
37 42 2.9 1035 4 US-09-834-859-1
38 42 2.9 1035 4 US-09-760-892-1
39 42 2.9 1035 4 US-09-834-856-1
40 42 2.9 1052 4 US-09-489-847-23
41 42 2.9 1174 2 US-08-872-437-1
42 42 2.9 1174 3 US-08-651-136C-11
43 42 2.9 1174 4 US-09-329-311A-11
44 42 2.9 1223 3 US-09-154-874-4
45 42 2.9 1223 4 US-08-931-668-4

```

ALIGNMENTS

```

RESULT 1
US-09-672-785-3
; Sequence 3, Application US/09672785
; Patent No. 6596927
; GENERAL INFORMATION:
; APPLICANT: Mizutani, Masako
; APPLICANT: Kasumi, Takaaki
; APPLICANT: Ayabe, Shin-ichi
; APPLICANT: Akashi, Tomoyoshi
; TITLE OF INVENTION: Genes Coding for Flavone Synthases
; FILE REFERENCE: 001560-383
; CURRENT APPLICATION NUMBER: US/09/672,785
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: PCT/JP00/04379
; PRIOR FILING DATE: 2000-01-30
; PRIOR APPLICATION NUMBER: PCT/JP00/00490
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: JP 11-205229
; PRIOR FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: JP 11-22427
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1730
; TYPE: DNA
; ORGANISM: Torenia hybrida
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence encoding a protein having an
; OTHER INFORMATION: activity to directly convert flavanone to flavone
US-09-672-785-3

```

```

Query Match 3.1%; Score 44; DB 4; Length 1730;
Best Local Similarity 100.0%; Pred. No. 7.4e-09;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1391 GATTTTAAAAA
DB 1680 GATTTTAAAAA

```

```

RESULT 2
US-08-481-190-15
; Sequence 15, Application US/08481190
; Patent No. 6160204
; GENERAL INFORMATION:
; APPLICANT: John C. Steffens
; TITLE OF INVENTION: Polyphenol Oxidase cDNA
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:

```

Wed Jul 21 09:11:32 2004

```

; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/481,190
; PRIOR APPLICATION NUMBER:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 203,533
; FILING DATE: 02-24-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: George M. Yahwak
; REGISTRATION NUMBER: 26,824
; REFERENCE/DOCKET NUMBER: UA 816 CIP
; TELEPHONE: (203)268-1951
; TELEFAX: (203)268-1951
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-481-190-15

Query Match 3.0%; Score 43; DB 3; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
Db 172 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 214

RESULT 3
PCT-US93-00869-15
; Sequence 15, Application PC/TUS9300869
; GENERAL INFORMATION:
; APPLICANT: John C. Steffens
; TITLE OF INVENTION: Polyphenol Oxidase cDNAs: Cloning
; TITLE OF INVENTION: and Applications
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00869
; FILING DATE: 19930129
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: George M. Yahwak
; REGISTRATION NUMBER: 26,824
; REFERENCE/DOCKET NUMBER: CRF D-1057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203)268-1951

; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,970A
; FILING DATE: 19920518
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 10 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB10
; CLONE: A9
; US-07-885-970A-13

; TELEFAX: (203)268-1951
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; PCT-US93-00869-15

Query Match 3.0%; Score 43; DB 5; Length 222;
Best Local Similarity 100.0%; Pred. No. 2.5e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
Db 172 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 214

RESULT 4
US-07-885-970A-13
; Sequence 13, Application US/07885970A
; Patent No. 5495070
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,970A
; FILING DATE: 19920518
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 10 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB10
; CLONE: A9
; US-07-885-970A-13
```


Query Match 3.0%; Score 43; DB 1; Length 399;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1392 ATTTTAAAAA 1434
DB 345 ATTTTAAAAA 387

RESULT 5
US-08-298-687A-13
; Sequence 13, Application US/08298687A
; Patent No. 5521078
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701

COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,687A
; FILING DATE: 21-NOV-1990
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 10 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB10
; CLONE: A9
US-08-298-687A-13

Query Match 3.0%; Score 43; DB 1; Length 399;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1392 ATTTTAAAAA 1434
DB 345 ATTTTAAAAA 387

RESULT 6
US-08-530-797-12/C
; Sequence 12, Application US/08530797
; Patent No. 5597718
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; APPLICANT: Umbeck, Paul F.
; APPLICANT: Brill, Winston J.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
; TITLE OF INVENTION: FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O. Box 2113
; STREET: FIRST WISCONSIN PLAZA
; CITY: MADISON
; STATE: WISCONSIN
; COUNTRY: U.S.A.
; ZIP: 53701
COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,797
; FILING DATE: 20-SEP-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/617,239
; FILING DATE: 21-NOV-90
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-88
; ATTORNEY/AGENT INFORMATION:
; NAME: Nicholas J. Seay
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990245
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 10 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB10
; CLONE: A9
US-08-530-797-12

Query Match 3.0%; Score 43; DB 1; Length 399;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1392 ATTTTAAAAA 1434
DB 55 ATTTTAAAAA 13

RESULT 7
US-08-829-829-13
; Sequence 13, Application US/08298829
; Patent No. 5620882
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.

```

;
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady
; STREET: P.O. Box 2113, First Wisconsin Plaza
; CITY: Madison
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53701
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/298,829
; FILING DATE: 19-OCT-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/885,970
; FILING DATE: 18-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/617,239
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27,386
; TELEPHONE: (608) 283-2478
; TELEFAX: (608) 251-5139
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 10 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB10
; CLONE: A9
;
; US-08-298-829-13

```

```

Query Match 3.0%; Score 43; DB 1; Length 399;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1392 ATTTTAAAAA 1434
Db 345 ATTTTAAAAA 387

RESULT 8
US-08-787-335-12/c
; Sequence 12, Application US/08/787335
; Patent No. 5981834
; GENERAL INFORMATION:
; APPLICANT: John, Maliyakal E.
; APPLICANT: Umbeck, Paul F.
; APPLICANT: Brill, Winston J.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COTTON PLANTS
; TITLE OF INVENTION: FOR ALTERED FIBER
; NUMBER OF SEQUENCES: 18

```

```

;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles and Brady
; STREET: P.O. BOX 2113
; STREET: FIRST WISCONSIN PLAZA
; CITY: MADISON
; STATE: WISCONSIN
; COUNTRY: U.S.A.
; ZIP: 53701
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.50 inch, 800Kb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/787,335
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/530,797
; FILING DATE:
; APPLICATION NUMBER: US 07/253,243
; FILING DATE: 04-OCT-88
; ATTORNEY/AGENT INFORMATION:
; NAME: Nicholas J. Seay
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 1122990245
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; ORGANISM: Gossypium hirsutum
; STRAIN: Coker 312
; DEVELOPMENTAL STAGE: 10 day old fiber cells
; TISSUE TYPE: fiber cells
; IMMEDIATE SOURCE:
; LIBRARY: CKFB10
; CLONE: A9
;
; US-08-787-335-12

```

```

Query Match 3.0%; Score 43; DB 2; Length 399;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1392 ATTTTAAAAA 1434
Db 55 ATTTTAAAAA 13

RESULT 9
US-09-152-060-41
; Sequence 41, Application US/09152060
; Patent No. 6448230
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 28 Human Secreted Proteins
; FILE REFERENCE: PZ003P1.US
; CURRENT APPLICATION NUMBER: US/09/152,060
; CURRENT FILING DATE: 1998-09-11
; EARLIER APPLICATION NUMBER: PCT/US98/04858
; EARLIER FILING DATE: 1998-03-12
; EARLIER APPLICATION NUMBER: 60/040,762
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/040,710
; EARLIER FILING DATE: 1997-03-14
; EARLIER APPLICATION NUMBER: 60/050,934
; EARLIER FILING DATE: 1997-05-30
; EARLIER APPLICATION NUMBER: 60/048,100

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (192)..(803)
US-09-250-611-3

Query Match          3.0%; Score 43; DB 4; Length 1307;
Best Local Similarity 100.0%; Pred. No. 1.9e-08;
Matches 43; Conservative 0; Mismatches 0; Indels

QY 1392 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
   |||||||
DB 1263 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1305

RESULT 12
US-09-614-912-175
; Sequence 175, Application US/09614912
; Patent No. 6677502
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Rafalski, Antoni
; APPLICANT: Orozco, Buddy
; APPLICANT: Miao, Gou-Hau
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Lee, Jian Ming
; APPLICANT: Sakai, Hajime
; APPLICANT: Weng, Zude
; APPLICANT: Cai, Perry G
; APPLICANT: Anderson, Shawn
; TITLE OF INVENTION: Plant Metabolism Genes
; FILE REFERENCE: BB1378 US NA
; CURRENT APPLICATION NUMBER: US/09/614,912
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: 60/143,401
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/143,412
; PRIOR FILING DATE: 1999-07-12
; PRIOR APPLICATION NUMBER: 60/146,650
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 60/170,906
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: 60/172,959
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/172,946
; PRIOR FILING DATE: 1999-12-21
; NUMBER OF SEQ ID NOS: 204
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 175
; LENGTH: 1919
; TYPE: DNA
; ORGANISM: Triticum aestivum
US-09-614-912-175

Query Match          3.0%; Score 43; DB 4; Length 1919;
Best Local Similarity 100.0%; Pred. No. 1.8e-08;
Matches 43; Conservative 0; Mismatches 0; Indels

QY 1392 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
   |||||||
DB 1876 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1918

RESULT 13
US-09-920-759-10
; Sequence 10, Application US/09920759
; Patent No. 6537811
; GENERAL INFORMATION:
; APPLICANT: Brenda F. Baker
; APPLICANT: Susan M. Freier
; TITLE OF INVENTION: ANTISENSE MODULATION OF SAP-1 EXPRESSION
; FILE REFERENCE: RTS-0267

```

Search completed: July 21, 2004, 00:04:40
Job time : 129 secs

1392 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1434
|||||
1921 ATTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1963

Query Match	3.0%;	Score 43;	DB 4;	Length 2202;
Best Local Similarity	100.0%;	Pred. No. 1.8e-08;		
Matches 43:	Conservative	0;	Mismatches 0;	Indels 0;
	Gaps	0;		

1392 ATTTTAAAAAAAAAAAAAAAAAAAAA 1434
|||||
2148 ATTTTAAAAAAAAAAAAAAAAAAAAA 2190

```

RESULT 15
US-09-439-923-1
// Sequence 1, Application US/09439923
// Patent No. 6426208
// GENERAL INFORMATION:
// APPLICANT: Emil D. Kakkis
// APPLICANT: Becky Tanamachi
// TITLE OF INVENTION: Recombinant Alpha-L-iduronidase, Methods
// TITLE OF INVENTION: for Producing and Purifying the Same and
// TITLE OF INVENTION: Treating Diseases Caused by Deficiencies Thereof
// TITLE OF INVENTION:
// FILE REFERENCE: 080000511US00
// CURRENT APPLICATION NUMBER: US/09/439,923
// CURRENT FILING DATE: 1999-11-12
// NUMBER OF SEQ ID NOS: 2
// SOFTWARE: FastSeq for Windows Version 3.0
// SEQ ID NO 1
// LENGTH: 6200
// TYPE: DNA
// ORGANISM: Homo sapiens
// FEATURE:

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2004, 23:04:30 ; Search time 677 Seconds
(without alignments)
10333.750 Million cell updates/sec

Title: US-10-799-747-12
Perfect score: 1434
Sequence: 1 cattaaactctttttatcgg.....aaaaaaaaaaaaaaaaaaaa 1434

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3191023 seqs, 2439312756 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6382046

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1434	100.0	1434	15	US-10-799-730-12
2	345	24.1	448	9	US-09-880-107-1385
3	47	3.3	8281	13	US-10-342-887-309
4	47	3.3	8281	13	US-10-172-118-309
5	45	3.1	493	10	US-09-770-961-611
6	45	3.1	1231	9	US-09-891-126-3
7	45	3.1	1231	9	US-10-266-745-3
8	45	3.1	2552	16	US-10-264-237-477
9	44	3.1	1730	16	US-10-411-115-3
10	43	3.0	193	10	US-09-814-353-18984
11	43	3.0	240	13	US-10-424-599-44257
12	43	3.0	282	10	US-09-814-353-17457
13	43	3.0	294	10	US-09-814-353-5106
14	43	3.0	294	10	US-09-814-353-11398

c	15	43	3.0	306	16	US-10-131-827-8716	Sequence 8716, Ap
c	16	43	3.0	337	9	US-09-770-791-889	Sequence 889, App
c	17	43	3.0	338	13	US-10-424-599-53632	Sequence 53632, A
c	18	43	3.0	341	17	US-10-021-323-451	Sequence 451, App
c	19	43	3.0	342	17	US-10-021-323-9783	Sequence 9783, Ap
c	20	43	3.0	344	13	US-10-424-599-35199	Sequence 35199, A
c	21	43	3.0	361	13	US-10-424-599-12247	Sequence 12247, A
c	22	43	3.0	374	13	US-10-424-599-50183	Sequence 50183, A
c	23	43	3.0	382	13	US-10-424-599-67598	Sequence 67598, A
c	24	43	3.0	382	13	US-10-424-599-82042	Sequence 82042, A
c	25	43	3.0	390	17	US-10-437-963-95670	Sequence 95670, A
c	26	43	3.0	398	17	US-10-021-323-7863	Sequence 7863, Ap
c	27	43	3.0	402	10	US-09-918-995-16715	Sequence 5673, Ap
c	28	43	3.0	403	13	US-10-424-599-5673	Sequence 16715, A
c	29	43	3.0	408	17	US-10-437-963-82810	Sequence 82810, A
c	30	43	3.0	415	17	US-10-424-599-5972	Sequence 5972, Ap
c	31	43	3.0	432	13	US-10-424-599-107176	Sequence 107176, A
c	32	43	3.0	433	13	US-10-424-599-122494	Sequence 122494, A
c	33	43	3.0	455	13	US-10-424-599-105065	Sequence 105065, A
c	34	43	3.0	457	17	US-10-424-599-113203	Sequence 113203, A
c	35	43	3.0	459	10	US-09-814-353-17965	Sequence 50630, A
c	36	43	3.0	480	16	US-10-131-827-8192	Sequence 17965, A
c	37	43	3.0	489	17	US-10-021-323-17555	Sequence 17555, A
c	38	43	3.0	492	17	US-10-437-963-29142	Sequence 29142, A
c	39	43	3.0	502	13	US-10-424-599-33446	Sequence 33446, A
c	40	43	3.0	510	17	US-10-437-963-26905	Sequence 26905, A
c	41	43	3.0	522	17	US-10-021-323-1498	Sequence 1498, Ap
c	42	43	3.0	526	17	US-10-021-323-14764	Sequence 14764, A
c	43	43	3.0	528	13	US-10-424-599-46393	Sequence 46393, A
c	44	43	3.0	528	16	US-10-131-827-8186	Sequence 8186, Ap
c	45	43	3.0	528	16	US-10-131-827-8186	Sequence 8186, Ap

ALIGNMENTS

RESULT 1
US-10-195-730-12
; Sequence 12, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 1434
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-195-730-12

Query Match	100.0%	Score 1434;	DB 15;	Length 1434;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1434;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	CATTAAACTCTTTTATCGGAATAGTAGATATTTCAATGTCTACTCATTCATGTGA	60	
Db	1	CATTAAACTCTTTTATCGGAATAGTAGATATTTCAATGTCTACTCATTCATGTGA	60	
QY	61	TTTGAGCTGCACAGTTATTTTGTCTACGACAGATTTTATTTTATTTTAAAGTCAGTGC	120	
Db	61	TTTGAGCTGCACAGTTATTTTGTCTACGACAGATTTTATTTTATTTTAAAGTCAGTGC	120	
QY	121	AAAATTATGATAGATATATACTAATAAATACAAAGTACAAAGTCAGAGCTGT	180	

Db	121	AAAAATATGATAGGATATACATAATAATAACAAGTAATAACAAAAGTCAAGACAGTGT	180
Qy	181	CTAAATATAAATTTCTGGGTTCTTAAATAATTTTAAATTTATCTTGAATAGTTTTCT	240
Db	181	CTAAATATAAATTTCTGGGTTCTTAAATAATTTTAAATTTATCTTGAATAGTTTTCT	240
Qy	241	TAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATATAA	300
Db	241	TAGATTAATCTCAGGATATGAGAAAGTCAATTAAGTGTGAGTAAAGTTAGTATCATATAA	300
Qy	301	CAAAATGCTATTAATGAMGAGTGTATATATACAGAAATTTATCAGGCATTAACCAAGTC	360
Db	301	CAAAATGCTATTAATGAMGAGTGTATATATACAGAAATTTATCAGGCATTAACCAAGTC	360
Qy	361	TAGGCATATAGGAATGCAGCACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTGTA	420
Db	361	TAGGCATATAGGAATGCAGCACTCAGAAATGGTTTCAATGTAGTAGTTGATGCTTGTA	420
Qy	421	AGGTAGGGAGCTTATTCAGACATAGTAGATAGTTTCTTAATGCTGTSTCAATTTGCTGG	480
Db	421	AGGTAGGGAGCTTATTCAGACATAGTAGATAGTTTCTTAATGCTGTSTCAATTTGCTGG	480
Qy	481	CTTTTGGCTACCTGACTTCCSCAATATGCGAGCCCAATTCAGTCTTGAGTTTCTTCTCT	540
Db	481	CTTTTGGCTACCTGACTTCCSCAATATGCGAGCCCAATTCAGTCTTGAGTTTCTTCTCT	540
Qy	541	GGACACCTTATGCTGGAATCAATGAGCGAGCTGATTCATTTGGTGAATTTGGGTAGAAA	600
Db	541	GGACACCTTATGCTGGAATCAATGAGCGAGCTGATTCATTTGGTGAATTTGGGTAGAAA	600
Qy	601	GCAGTATGTTTGTGTCATTAAGATGTAGGTATATAGATAGTTTACGCTTTAAGTGTAT	660
Db	601	GCAGTATGTTTGTGTCATTAAGATGTAGGTATATAGATAGTTTACGCTTTAAGTGTAT	660
Qy	661	GTTTTATATCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCACCTCTCCCCGAGC	720
Db	661	GTTTTATATCTTTAAATAAGAAATATAACCTTTTAAAGCTATTCACCTCTCCCCGAGC	720
Qy	721	CTATCTCAAACTGGTGAATATATGAGAGATCTTTGAAGAAGTAAATAAAACCTTCACT	780
Db	721	CTATCTCAAACTGGTGAATATATGAGAGATCTTTGAAGAAGTAAATAAAACCTTCACT	780
Qy	781	GCTCCACTCCAGGTGAATCCGCCCCACTCCACTGACCTAGTAGAATTTGTAATTAATAC	840
Db	781	GCTCCACTCCAGGTGAATCCGCCCCACTCCACTGACCTAGTAGAATTTGTAATTAATAC	840
Qy	841	TTACCTTCTATTTCTGAATCAGTTGTGAATGTTGCTTATGTTTCAGAGTTTAAAGAAC	900
Db	841	TTACCTTCTATTTCTGAATCAGTTGTGAATGTTGCTTATGTTTCAGAGTTTAAAGAAC	900
Qy	901	CTCMGTGAATTCATTTTAAATCTGCTATTTCTGAGAAGCATTTGAATGAATTTCTTAACA	960
Db	901	CTCMGTGAATTCATTTTAAATCTGCTATTTCTGAGAAGCATTTGAATGAATTTCTTAACA	960
Qy	961	AGAAGACTCATCTGAGCTGTTTGTGCTATGAGCCCCATAGGGTTCTGTCGCTTA	1020
Db	961	AGAAGACTCATCTGAGCTGTTTGTGCTATGAGCCCCATAGGGTTCTGTCGCTTA	1020
Qy	1021	GCATTTAAACAAATAGGTTTATAGTAAAGCCAATGTATTAATTTTTCATCGAGGAGG	1080
Db	1021	GCATTTAAACAAATAGGTTTATAGTAAAGCCAATGTATTAATTTTTCATCGAGGAGG	1080
Qy	1081	GCTTTTAAATTTGTGCTCTTTTTCATATTTTATCATATTCATATTTGTTGTTGAACTG	1140
Db	1081	GCTTTTAAATTTGTGCTCTTTTTCATATTTTATCATATTTGTTGTTGTTGAACTG	1140
Qy	1141	CTTTTATAGGAGAT	1200
Db	1141	CTTTTATAGGAGAT	1200
Qy	1201	GGATAATTTAATTTACGTGCTTCTGTTATTCAGATAAAGAGAGACTACGCTGCATA	1260

Db	1201	GGATAATTTAATTTACGTGCTTCTGTTATTCAGATAAAGAGAGACTACGCTGCATA	1260
Qy	1261	TTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTCTTAAGATTTTCAAAGGAAATA	1320
Db	1261	TTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTCTTAAGATTTTCAAAGGAAATA	1320
Qy	1321	GTGTAATTTGAGAAATCATAAACCACTGCTTAACCTTGGTAAACAAACCTGTTCTTAATAA	1380
Db	1321	GTGTAATTTGAGAAATCATAAACCACTGCTTAACCTTGGTAAACAAACCTGTTCTTAATAA	1380
Qy	1381	AGTATTAAATGATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1434
Db	1381	AGTATTAAATGATTTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	1434

RESULT 2

US-09-880-107-1385/c

; Sequence 1385, Application US/09880107

; Patent No. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

; APPLICANT: Vockley, Joseph G.

; APPLICANT: Scherf, Uwe

; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer

; FILE REFERENCE: 44921-5028-WO

; CURRENT APPLICATION NUMBER: US/09/880,107

; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: US 60/211,379

; PRIOR FILING DATE: 2000-06-14

; PRIOR APPLICATION NUMBER: US 60/237,054

; PRIOR FILING DATE: 2000-10-02

; NUMBER OF SEQ ID NOS: 3950

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1385

; LENGTH: 448

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AA491000

US-09-880-107-1385

Query Match 24.1%; Score 345; DB 9; Length 448;

Best Local Similarity 100.0%; Pred. No. 2.1e-150;

Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1000	CCCATAGGGTTCTGCTTAGCATTAACAAATAAGGTTTATAGGTAAAGCCCAATGTAT	1059
Db	414	CCCATAGGGTTCTGCTTAGCATTAACAAATAAGGTTTATAGGTAAAGCCCAATGTAT	355
Qy	1060	TAAATTTTTTTTGCATGGAGGGCTTTAAAAATTTGCTCTTTTTCATATTTTATTCATAT	1119
Db	354	TAAATTTTTTTTGCATGGAGGGCTTTAAAAATTTGCTCTTTTTCATATTTTATTCATAT	295
Qy	1120	TCAATTTATAGTTTGTAACTGCTTTTATAGGAGATAATATATATATATATATATATAT	1179
Db	294	TCAATTTATAGTTTGTAACTGCTTTTATAGGAGATAATATATATATATATATATATAT	235
Qy	1180	GGGGGAATAATTTGTCAGAGGATAATTTAAATTTACGCTTCTGTTTATTCAGAAATAA	1239
Db	234	GGGGGAATAATTTGTCAGAGGATAATTTAAATTTACGCTTCTGTTTATTCAGAAATAA	175
Qy	1240	GAGAGAAGACTACGCTGCATATTCAGAGTTGTACCTTAAACATTTGGTGAACATTTTTC	1299
Db	174	GAGAGAAGACTACGCTGCATATTCAGAGTTGTACCTTAAACATTTGGTGAACATTTTTC	115
Qy	1300	TAAATTTTCAAAAGGAATAATGTAATTTGAGAAATCATATACCA	1344
Db	114	TAAATTTTCAAAAGGAATAATGTAATTTGAGAAATCATATACCA	70

RESULT 3

US-10-342-887-309

```
; Sequence 309, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 309
; LENGTH: 8281
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-309

Query Match          3.3%; Score 47; DB 13; Length 8281;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AATGATTTTAAAAA 1434
Db 8219 AATGATTTTAAAAA 1434

RESULT 4
US-10-172-118-309
; Sequence 309, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 309
; LENGTH: 8281
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AL133623
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-309

Query Match          3.3%; Score 47; DB 13; Length 8281;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AATGATTTTAAAAA 1434
Db 8219 AATGATTTTAAAAA 1434

; Sequence 309, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 309
; LENGTH: 8281
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-309

Query Match          3.3%; Score 47; DB 13; Length 8281;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AATGATTTTAAAAA 1434
Db 8219 AATGATTTTAAAAA 1434

; Sequence 309, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 309
; LENGTH: 8281
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AL133623
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-309

Query Match          3.3%; Score 47; DB 13; Length 8281;
Best Local Similarity 100.0%; Pred. No. 3.2e-11;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1388 AATGATTTTAAAAA 1434
Db 8219 AATGATTTTAAAAA 1434
```

```
RESULT 5
US-09-770-961-611/c
; Sequence 611, Application US/09770961
; Publication No. US20030115639A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krickler, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; FILE REFERENCE: 2026 (PARA-015PRV)
; CURRENT APPLICATION NUMBER: US/09/770,961
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,466
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 611
; LENGTH: 493
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-961-611

Query Match          3.1%; Score 45; DB 10; Length 493;
Best Local Similarity 100.0%; Pred. No. 2.1e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA 1434
Db 52 TGATTTTAAAAA 1434

RESULT 6
US-09-891-126-3
; Sequence 3, Application US/09891126
; Patent No. US20020072596A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Transferrin Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT035PI
; CURRENT APPLICATION NUMBER: US/09/891,126
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: PCT/US00/34769
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/171,595
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-891-126-3

Query Match          3.1%; Score 45; DB 9; Length 1231;
```



```
Best Local Similarity 100.0%; Pred. No. 2.3e-10; Indels 0; Gaps 0;
Matches 45; Conservative 0; Mismatches 0;

QY 1390 TGATTTTAAAAA 1434
Db 1147 TGATTTTAAAAA 1191

RESULT 7
US-10-266-745-3
; Sequence 3, Application US/10266745
; Publication No. US20030149256A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Transferrin Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT035P1
; CURRENT APPLICATION NUMBER: US/10/266,745
; CURRENT FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: US/09/891,126
; PRIOR FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: PCT/US00/34769
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/171,595
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1231
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-266-745-3

Query Match 3.1%; Score 45; DB 15; Length 1231;
Best Local Similarity 100.0%; Pred. No. 2.3e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA 1434
Db 1147 TGATTTTAAAAA 1191

RESULT 8
US-10-264-237-477
; Sequence 477, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 477
; LENGTH: 2552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-237-477

Query Match 3.1%; Score 45; DB 16; Length 2552;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA 1434
Db 2504 TGATTTTAAAAA 2548

RESULT 9
US-10-264-237-477
; Sequence 477, Application US/10264237
; Publication No. US20040009491A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA131P1
; CURRENT APPLICATION NUMBER: US/10/264,237
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/16450
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2876
; SOFTWARE: Patentin Ver. 3.1
; SEQ ID NO 477
; LENGTH: 2552
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-264-237-477

Query Match 3.1%; Score 45; DB 16; Length 2552;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1390 TGATTTTAAAAA 1434
Db 2504 TGATTTTAAAAA 2548

RESULT 10
US-09-814-353-18984/c
; Sequence 18984, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18984
; LENGTH: 193
; TYPE: DNA
```

```
US-10-411-115-3
; Sequence 3, Application US/10411115
; Publication No. US20040003431A1
; GENERAL INFORMATION:
; APPLICANT: Mizutani, Masako
; APPLICANT: Kasumi, Takaaki
; APPLICANT: Ayabe, Shin-ichi
; APPLICANT: Akashi, Tomoyoshi
; TITLE OF INVENTION: Genes Coding for Flavone Synthases
; FILE REFERENCE: 001560-383
; CURRENT APPLICATION NUMBER: US/10/411,115
; CURRENT FILING DATE: 2003-04-11
; PRIOR APPLICATION NUMBER: US/09/672,785
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: PCT/JP00/04379
; PRIOR FILING DATE: 2000-01-30
; PRIOR APPLICATION NUMBER: PCT/JP00/00490
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: JP 11-205229
; PRIOR FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: JP 11-22427
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 1730
; TYPE: DNA
; ORGANISM: Torenia hybrida
; FEATURE:
; OTHER INFORMATION: Nucleotide sequence encoding a protein having an
; OTHER INFORMATION: activity to directly convert flavanone to flavone
US-10-411-115-3

Query Match 3.1%; Score 44; DB 16; Length 1730;
Best Local Similarity 100.0%; Pred. No. 7e-10;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1391 GATTTTAAAAA 1434
Db 1680 GATTTTAAAAA 1723

RESULT 10
US-09-814-353-18984/c
; Sequence 18984, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18984
; LENGTH: 193
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-09-814-353-18984

Query Match          3.0%; Score 43; DB 10; Length 193;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 3.0%; Score 43; DB 10; Length 282;
      ||||| 100.0%; Pred. No. 1.7e-09;
      ||||| 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 170 ATTTTAAAAA 3.0%; Score 43; DB 10; Length 282;
      ||||| 100.0%; Pred. No. 1.7e-09;
      ||||| 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11
US-10-424-599-44257
; Sequence 44257, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 44257
; LENGTH: 240
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_139963C.1
US-10-424-599-44257

Query Match          3.0%; Score 43; DB 13; Length 240;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 3.0%; Score 43; DB 13; Length 240;
      ||||| 100.0%; Pred. No. 1.7e-09;
      ||||| 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 27 ATTTTAAAAA 3.0%; Score 43; DB 13; Length 240;
      ||||| 100.0%; Pred. No. 1.7e-09;
      ||||| 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
US-09-814-353-17457/c
; Sequence 17457, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17457
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 126, 127, 128, 142, 143, 144, 145, 146, 147, 148, 149, 162,
; LOCATION: 174, 176, 177, 178, 181, 199, 206, 213, 223, 227, 228, 232,
; LOCATION: 235, 255
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-17457/c

Query Match          3.0%; Score 43; DB 10; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 3.0%; Score 43; DB 10; Length 294;
      ||||| 100.0%; Pred. No. 1.7e-09;
      ||||| 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 124 ATTTTAAAAA 3.0%; Score 43; DB 10; Length 294;
      ||||| 100.0%; Pred. No. 1.7e-09;
      ||||| 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
US-09-814-353-11398/c
; Sequence 11398, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
```

```
; LENGTH: 282
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-814-353-17457

Query Match          3.0%; Score 43; DB 10; Length 282;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 3.0%; Score 43; DB 10; Length 282;
      ||||| 100.0%; Pred. No. 1.7e-09;
      ||||| 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 217 ATTTTAAAAA 3.0%; Score 43; DB 10; Length 282;
      ||||| 100.0%; Pred. No. 1.7e-09;
      ||||| 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 13
US-09-814-353-5106/c
; Sequence 5106, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5106
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 126, 127, 128, 142, 143, 144, 145, 146, 147, 148, 149, 162,
; LOCATION: 174, 176, 177, 178, 181, 199, 206, 213, 223, 227, 228, 232,
; LOCATION: 235, 255
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-5106

Query Match          3.0%; Score 43; DB 10; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1392 ATTTTAAAAA 3.0%; Score 43; DB 10; Length 294;
      ||||| 100.0%; Pred. No. 1.7e-09;
      ||||| 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 124 ATTTTAAAAA 3.0%; Score 43; DB 10; Length 294;
      ||||| 100.0%; Pred. No. 1.7e-09;
      ||||| 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 14
US-09-814-353-11398/c
; Sequence 11398, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
```

Search completed: July 21, 2004, 01:37:12
Job time : 679 secs

```

; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11398
; LENGTH: 294
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 126, 127, 128, 142, 143, 144, 145, 146, 147, 148, 149, 162,
; LOCATION: 174, 176, 177, 178, 181, 199, 206, 213, 223, 227, 228, 232,
; LOCATION: 235, 255
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-11398

Query Match 3.0%; Score 43; DB 10; Length 294;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1392 ATTTTAAAAA 1434
Db 124 ATTTTAAAAA 82

RESULT 15
US-10-131-827-8716/c
; Sequence 8716, Application US/10131827
; Publication No. US20040009479A1
; GENERAL INFORMATION:
; APPLICANT: Wohlgemuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: LV, Ngoc
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 56661200120
; CURRENT APPLICATION NUMBER: US/10/131,827
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8716
; LENGTH: 306
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-827-8716

Query Match 3.0%; Score 43; DB 16; Length 306;
Best Local Similarity 100.0%; Pred. No. 1.7e-09;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1392 ATTTTAAAAA 1434
Db 67 ATTTTAAAAA 25

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: July 20, 2004, 21:04:06 ; Search time 3468 Seconds
(without alignments)
12347.857 Million cell updates/sec

Title: US-10-799-747-12
Perfect score: 1434
Sequence: 1 cattaaactctttttatcgg.....aaaaaaaaaaaaaaaaaaaa 1434

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST*

- 1: em_estba:**
- 2: em_esthum:**
- 3: em_estin:**
- 4: em_estnu:**
- 5: em_estov:**
- 6: em_estpl:**
- 7: em_estro:**
- 8: em_htc:**
- 9: gb_estl:**
- 10: gb_est2:**
- 11: gb_htc:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estcom:**
- 17: em_gss_hum:**
- 18: em_gss_inv:**
- 19: em_gss_pln:**
- 20: em_gss_vrt:**
- 21: em_gss_fun:**
- 22: em_gss_man:**
- 23: em_gss_mus:**
- 24: em_gss_pro:**
- 25: em_gss_rod:**
- 26: em_gss_pbg:**
- 27: em_gss_vrl:**
- 28: gb_gsl1:**
- 29: gb_gsl2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	448	31.2	814	14	CB853053	CB853053 UI-CF-FNO
C 2	444	31.0	444	9	AA136080	AA136080 2k90d03.s
C 3	421	29.4	551	12	B1259796	B1259796 602971440
C 4	414	28.9	556	10	AW500190	AW500190 UI-HF-BNO

C 5	401	28.0	588	10	BF431622	BF431622 7016e08.x
C 6	400	27.9	438	10	AW304923	AW304923 xv97b07.x
C 7	395	27.5	395	9	AA928811	AA928811 ot85b03.s
C 8	364	25.4	400	9	AW002842	AW002842 wr56h11.x
C 9	346	24.1	1201	13	EX441923	EX441923 BX441923
C 10	345	24.1	448	9	AA491000	AA491000 aa52g11.s
C 11	336	23.4	785	12	EG484396	EG484396 602505037
C 12	333	21.8	941	13	EX452125	EX452125 BX452125
C 13	303	21.1	419	10	AW189910	AW189910 xl10a04.x
C 14	286	19.9	568	9	AL035942	AL035942 DKF2p564B
C 15	284	19.8	557	14	CB132625	CB132625 K-EST0183
C 16	276	19.2	280	9	AA679646	AA679646 ah10c06.s
C 17	254	17.7	493	9	AW026264	AW026264 wv10h02.x
C 18	240	16.7	394	9	AA747222	AA747222 dx76a10.s
C 19	227	15.8	621	9	AL035941	AL035941 DKF2p564B
C 20	227	15.8	688	12	BM728856	BM728856 UI-E-RO1-
C 21	203	14.2	575	12	BM722991	BM722991 UI-E-RO1-
C 22	200	13.9	253	9	AA953459	AA953459 on69g09.s
C 23	181	12.6	503	10	AW889139	AW889139 QV4-NT002
C 24	170	11.9	765	12	BI914473	BI914473 603182264
C 25	149	10.4	957	12	EG403642	EG403642 602419361
C 26	105	7.3	1037	13	EX413464	EX413464 BX413464
C 27	104	7.3	336	12	BI025209	BI025209 CM4-NT028
C 28	82	5.7	634	13	EX506904	EX506904 DKF2p779F
C 29	77	5.4	240	10	BF910533	BF910533 CM4-TT004
C 30	60	4.2	1117	12	BM466406	BM466406 AGENCOURT
C 31	48	3.3	419	10	AW189910	AW189910 xl10a04.x
C 32	47	3.3	229	12	EG736035	EG736035 tk46h07.y
C 33	47	3.3	229	12	EG737212	EG737212 tk68c11.y
C 34	47	3.3	427	10	BE031403	BE031403 129955 MA
C 35	46	3.2	231	14	CF350700	CF350700 rl54b08.y
C 36	46	3.2	281	12	BI679838	BI679838 457100 MA
C 37	46	3.2	386	10	AW733394	AW733394 SK71f02.y
C 38	46	3.2	423	12	BI774735	BI774735 466945 WA
C 39	46	3.2	439	14	CF602204	CF602204 tac40f03
C 40	46	3.2	488	10	BE723220	BE723220 192661 MA
C 41	46	3.2	868	14	CK160202	CK160202 EGAS04176
C 42	45	3.1	102	9	AI447690	AI447690 mr30h12.x
C 43	45	3.1	116	14	CD520351	CD520351 AGENCOURT
C 44	45	3.1	134	9	AI499325	AI499325 t010c06.x
C 45	45	3.1	162	13	BQ189083	BQ189083 UI-E-EJ1-

ALIGNMENTS

RESULT 1
CB853053/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

CB853053 814 bp mRNA linear EST 22-APR-2003
UI-CF-FNO-afg-e-20-0-UI-s1 UI-CF-FNO Homo sapiens cDNA clone
UI-CF-FNO-afg-e-20-0-UI 3', mRNA sequence.
CB853053
CB853053
CB853053.1 GI:30048171
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 814)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa

cdNA library preparation: Dr. M. Bento Soares, University of Iowa
cdNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
The following repetitive elements were found in this cdNA
sequence: 1-42, >AT rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES Location/Qualifiers
Source
1..814
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-FNO-afg-e-20-0-UI"
/tissue_type="Human Lung Epithelial cells"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-FNO"
/note="Organ: Lung; Vector: p7T3-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-FNO is a subcloned cdNA library derived from two
normalized Human lung epithelial cell libraries (BN1 and
DU1) the library was subcloned according to according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. For additional information, contact:
bento-soares@uiowa.edu
TAG_SEQ=None found"

ORIGIN
Query Match 31.2%; Score 448; DB 14; Length 814;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 718; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 693 TTTAAGCTATTCCACCTCTCCCTCCAGCTATCTCAAACTGGTGGAAATATATGAGAGAT 752
Db |||||
QY 721 TTTAAGCTATTCCACCTCTCCCTCCAGCTATCTCAAACTGGTGGAAATATATGAGAGAT 662
Db |||||
QY 753 CTTGAAAGAGTAAATAAACCCTTCACCTGCTCCACTCAGGTGAATCGGCCACTCCAC 812
Db |||||
QY 661 CTTGAAAGAGTAAATAAACCCTTCACCTGCTCCACTCAGGTGAATCGGCCACTCCAC 602
QY 813 TGACCTAGTAGAATTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 872
Db |||||
QY 601 TGACCTAGTAGAATTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 542
QY 873 GTTCCTTATGCTCAGAGGTGTTAAGAACTCTGCTGAATTCATTTTTTAAATCTGCTATT 932
Db |||||
QY 541 GTTCCTTATGCTCAGAGGTGTTAAGAACTCTGCTGAATTCATTTTTTAAATCTGCTATT 482
QY 933 CTGGAAGCAATGAATGAATTTCTTAAAGAGAGCTCATCTGTAGCTGTTGCTGACTCC 992
Db |||||
QY 481 CTGGAAGCAATGAATGAATTTCTTAAAGAGAGCTCATCTGTAGCTGTTGCTGACTCC 422
QY 993 TATGAGCCCAATAGAGGTGCTGCTGCTAGCTTAAACAAATTAAGTTTATAGTAAAGCC 1052
Db |||||
QY 421 TATGAG-CCCAATAGAGGTGCTGCTGCTAGCTTAAACAAATTAAGTTTATAGTAAAGCC 363
QY 1053 AATGATTAATATTTTTTTGCTGAGGCGCTTTAAATTTGCTCTTTTTCATATTTTA 1112
Db |||||
QY 362 AATGATTAATATTTTTTTGCTGAGGCGCTTTAAATTTGCTCTTTTTCATATTTTA 303
QY 1113 TTCATATTCATATTTATGTTTGAATGCTTTTATAGGAGGATTAATATGTTTAAATTT 1172
Db |||||
QY 302 TTCATATTCATATTTATGTTTGAATGCTTTTATAGGAGGATTAATATGTTTAAATTT 243
QY 1173 AGTTTGGGGGGAATATTTGTCAGAGAGGATTAATTTAATTTACGTGCTCTGTTTATTTCA 1232
Db |||||
QY 242 AGTTTGGGGGGAATATTTGTCAGAGAGGATTAATTTAATTTACGTGCTCTGTTTATTTCA 183
QY 1233 GAATAAGAGAGAGAGACTACGCTGCATATTTCAAGAGTTGATCTTAACATTTGGTGAACA 1292
Db |||||
QY 182 GAATAAGAGAGAGAGACTACGCTGCATATTTCAAGAGTTGATCTTAACATTTGGTGAACA 123

QY 1293 TTTTCTTCTAAGATTTTCAAAAGGAATATGTAAATTGAGAAATCATACCACTGCTCTA 1352
Db |||||
QY 122 TTTTCTTCTAAGATTTTCAAAAGGAATATGTGTAAATTGAGAAATCATACCACTGCTCTA 63
QY 1353 ACTTGTGTAACAACTGTTCTTTAAATTAAGATTTTAAATGATTTTAAAAAATAAAAAA 1412
Db |||||
QY 62 ACTTGTGTAACAACTGTTCTTTAAATTAAGATTTTAAATGATTTTAAAAAATAAAAAA 3
QY 1413 AA 1414
Db |||
QY 2 AA 1

RESULT 2
AA136080/c
LOCUS
DEFINITION
IMAGE:490085 3', mRNA sequence.
ACCESSION
AA136080
VERSION
AA136080.1 GI:1697290
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 444)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,P.,
Trevasakis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 366.

FEATURES Location/Qualifiers
Source
1..444
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3804672"
/db_xref="taxon:9606"
/clone="IMAGE:490085"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="Soares_pregnant_uterus_NBHPU"
/note="Organ: uterus; Vector: p7T3-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cdNA was primed with a Not I -
oligo(dT) primer [5',
AACTGGAGATTCGGCGCCCTTTTTTTTTTTTTTTT 3'],
double-stranded cdNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

ORIGIN
Query Match 31.0%; Score 444; DB 9; Length 444;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 963 AAGACTCATCTCTAGCTGTTTCTGCTGCTCTCTATGAGCCCAATAAGGGTTCTGCTTAGC 1022
Db |||||
QY 444 AAGACTCATCTCTAGCTGTTTCTGCTGCTCTCTATGAGCCCAATAAGGGTTCTGCTTAGC 385


```

/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LT11)"
/clone_lib="NIH MGC 50"
/notes="Vector: pTV73-Pac; Site 1: NotI; Site 2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(3.5-4.4kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

ORIGIN
Query Match 28.9%; Score 414; DB 10; Length 556;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1000 CCCATAAGGGTCTGTGCTTAGCATTAACAAAATAAGGTTTATAGGTAAAGCCCAATGTAT 1059
Db 421 CCCATAAGGGTCTGTGCTTAGCATTAACAAAATAAGGTTTATAGGTAAAGCCCAATGTAT 362

QY 1060 TAATTTTTTTTGCATGAGGGCTTTAAATTTTGTGCTCTTTTCATATTTATTCATAT 1119
Db 361 TAATTTTTTTTGCATGAGGGCTTTAAATTTTGTGCTCTTTTCATATTTATTCATAT 302

QY 1120 TCAATTTATGTTTGAACCTGCTTTTAGGGAGATAATTTATATGTTTATTAATTTAGTTTG 1179
Db 301 TCAATTTATGTTTGAACCTGCTTTTAGGGAGATAATTTATATGTTTATTAATTTAGTTTG 242

QY 1180 GGGGGAATTAATTTGCAAGAGAGATAATTTAAATTTAGCTGCTCTGTTATTCAGATAAAA 1239
Db 241 GGGGGAATTAATTTGCAAGAGAGATAATTTAAATTTAGCTGCTCTGTTATTCAGATAAAA 182

QY 1240 GAGAGAAGACTACGCTGCATATTCAGAGTTGACCTTAACTTGGTGAACATTTTTTTC 1299
Db 181 GAGAGAAGACTACGCTGCATATTCAGAGTTGACCTTAACTTGGTGAACATTTTTTTC 122

QY 1300 TAAGATTTTCAAAGGAATATGTAAATTCAGAAATCATAACCACTGTCTTAACCTTGGT 1359
Db 121 TAAGATTTTCAAAGGAATATGTAAATTCAGAAATCATAACCACTGTCTTAACCTTGGT 62

QY 1360 AAACAAACTGTTCTTAAATAAGTATTTAATGATTTTAAAAAATAAAAAA 1413
Db 61 AAACAAACTGTTCTTAAATAAGTATTTAATGATTTTAAAAAATAAAAAA 8

RESULT 5
BF431622/c
LOCUS 7016e08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3574359 3',
DEFINITION mRNA sequence.
ACCESSION BF431622
VERSION BF431622.1 GI:11443736
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1. (bases 1 to 588)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Christopher Mokaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lemmon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 469.

FEATURES
source
Location/Qualifiers
1. 588
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3574359"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Kid11"
/notes="Organ: Kidney; Vector: pTV73D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."

ORIGIN
Query Match 28.0%; Score 401; DB 10; Length 588;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1000 CCCATAAGGGTCTGTGCTTAGCATTAACAAAATAAGGTTTATAGGTAAAGCCCAATGTAT 1059
Db 401 CCCATAAGGGTCTGTGCTTAGCATTAACAAAATAAGGTTTATAGGTAAAGCCCAATGTAT 342

QY 1060 TAATTTTTTTTGCATGAGGGCTTTAAATTTTGTGCTCTTTTCATATTTATTCATAT 1119
Db 341 TAATTTTTTTTGCATGAGGGCTTTAAATTTTGTGCTCTTTTCATATTTATTCATAT 282

QY 1120 TCAATTTATGTTTGAACCTGCTTTTAGGGAGATAATTTATATGTTTATTAATTTAGTTTG 1179
Db 281 TCAATTTATGTTTGAACCTGCTTTTAGGGAGATAATTTATATGTTTATTAATTTAGTTTG 222

QY 1180 GGGGGAATTAATTTGCAAGAGAGATAATTTAAATTTAGCTGCTCTGTTATTCAGATAAAA 1239
Db 221 GGGGGAATTAATTTGCAAGAGAGATAATTTAAATTTAGCTGCTCTGTTATTCAGATAAAA 162

QY 1240 GAGAGAAGACTACGCTGCATATTCAGAGTTGACCTTAACTTGGTGAACATTTTTTTC 1299
Db 161 GAGAGAAGACTACGCTGCATATTCAGAGTTGACCTTAACTTGGTGAACATTTTTTTC 102

QY 1300 TAAGATTTTCAAAGGAATATGTAAATTCAGAAATCATAACCACTGTCTTAACCTTGGT 1359
Db 101 TAAGATTTTCAAAGGAATATGTAAATTCAGAAATCATAACCACTGTCTTAACCTTGGT 42

QY 1360 AAACAAACTGTTCTTAAATAAGTATTTAATGATTTTAAAAA 1400
Db 41 AAACAAACTGTTCTTAAATAAGTATTTAATGATTTTAAAAA 1

RESULT 6
AW304923/c
LOCUS xv97b07.x1 NCI_CGAP_Brn53 Homo sapiens cDNA clone IMAGE:2826421 3',
DEFINITION mRNA sequence.
ACCESSION AW304923
VERSION AW304923.1 GI:6717126
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1. (bases 1 to 438)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGAP), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov

```


Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center
 Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyt not found
 Seq primer: -40UP from Gibco
 High quality sequence stop: 410.

FEATURES

Location/Qualifiers
 1. 438
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2826421"
 /tissue_types="three pooled meningiomas"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP Brn53"
 /notes="Organ: Brain; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies."

ORIGIN

Query Match 27.9%; Score 400; DB 10; Length 438;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1000 CCCTAAGGGTCTGTGCTAGCATTAACAAATAAGGTTTATAGGTAAGCCCAATGTAT 1059
 Db 400 CCCTAAGGGTCTGTGCTAGCATTAACAAATAAGGTTTATAGGTAAGCCCAATGTAT 341

QY 1060 TAAATTTTTTTCATGAGGGCTTTAAATTTGTGCTCTTTTCATATTTTTCATAT 1119
 Db 340 TAAATTTTTTTCATGAGGGCTTTAAATTTGTGCTCTTTTCATATTTTTCATAT 281

QY 1120 TCAATTTATGTTTGAACGCTTTTGGGAGATAATTATATGTTATAAATTTAGTTTG 1179
 Db 280 TCAATTTATGTTTGAACGCTTTTGGGAGATAATTATATGTTATAAATTTAGTTTG 221

QY 1180 GGGGGAATAATTGTGCAAGAGGATAATTAAATTTAGCTCTCTCTGTTATTCAGATAAA 1239
 Db 220 GGGGGAATAATTGTGCAAGAGGATAATTAAATTTAGCTCTCTCTGTTATTCAGATAAA 161

QY 1240 GAGAGAAGACTACGCTGCATTTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTTC 1299
 Db 160 GAGAGAAGACTACGCTGCATTTCAAGAGTTGTACCTTAACATTTGGTGAACATTTTTC 101

QY 1300 TAAGATTTTCAAGAAGATAATGTAAATTGAGAAATCATACCACTGCTCTTAACCTTGGT 1359
 Db 100 TAAGATTTTCAAGAAGATAATGTAAATTGAGAAATCATACCACTGCTCTTAACCTTGGT 41

QY 1360 AAACAACCTGTTCTTAAATAAGTATTTTAATGATTTTAA 1399
 Db 40 AAACAACCTGTTCTTAAATAAGTATTTTAATGATTTTAA 1

RESULT 7

AA92811/c
 LOCUS
 DEFINITION
 IMAGE:1623533 3', mRNA sequence.
 AA92811
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 395)
 AUTHORS
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE

JOURNAL

COMMENT

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 500 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 375.

FEATURES

source

Location/Qualifiers
 1. 395
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1623533"
 /dev_stage="8-9 weeks"
 /lab_host="DH10B"
 /clone_lib="Soares total fetus Nb2HF8 9w"
 /notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3']
 Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaudo."

ORIGIN

Query Match 27.5%; Score 395; DB 9; Length 395;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1011 TCTGTGCTTAGCATTAAACAAATAAGGTTTATAGTAAAGCCCAATGTTAATTTTTTTT 1070
 Db 395 TCTGTGCTTAGCATTAAACAAATAAGGTTTATAGTAAAGCCCAATGTTAATTTTTTTT 336

QY 1071 TGCATCGAGGGCTTAAATTTGCTCTTTTCATATTTTATTCATATTCAATTTATGG 1130
 Db 335 TGCATCGAGGGCTTAAATTTGCTCTTTTCATATTTTATTCATATTCAATTTATGG 276

QY 1131 TTGTAACTGCTTTTATAGGAGATAATTATATGTTATAAATTAGTTTGGGGGGAATAAT 1190
 Db 275 TTGTAACTGCTTTTATAGGAGATAATTATATGTTATAAATTAGTTTGGGGGGAATAAT 216

QY 1191 TGTGCAAGAGGATAATTTAATTTACGTCTTCTGTTATTTCAGAAATAAGAGAGAAGACT 1250
 Db 215 TGTGCAAGAGGATAATTTAATTTACGTCTTCTGTTATTTCAGAAATAAGAGAGAAGACT 156

QY 1251 AGCTGCATATTCAGAGTTGTACCTTAACATTTGGTGAACATTTTCTAAGATTTCAT 1310
 Db 155 AGCTGCATATTCAGAGTTGTACCTTAACATTTGGTGAACATTTTCTAAGATTTCAT 96

QY 1311 AAAGGAATATGTATAAATTGAGAAATCATACCACTGCTCAACTTGGTAAACAACACTGT 1370
 Db 95 AAAGGAATATGTATAAATTGAGAAATCATACCACTGCTCAACTTGGTAAACAACACTGT 36

QY 1371 TCTTAAATAAGTATTTAATGATTTTAAAAAATA 1405
 Db 35 TCTTAAATAAGTATTTAATGATTTTAAAAAATA 1

RESULT 8

AW002842/c
 LOCUS
 DEFINITION
 IMAGE:1623533 3', mRNA sequence.
 AW002842
 VERSION
 KEYWORDS
 SOURCE
 Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 400)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: Greg Lennon, Ph.D.
Cloning Distribution: Washington University Genome Sequencing Center
Cloned through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bhrp/image/image.html
Insert length: 1470 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 386.
Location/Qualifiers
FEATURES
source
1..400
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2491749"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI CGAP Utl"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

ORIGIN

Query Match 25.4%; Score 364; DB 9; Length 400;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 364; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1034 AAGGTTTATAGTAAAGCAATGATTAATTTTTCATGAGGAGGCTTTAAATTTG 1093
Db |||||||
Qy 364 AAGGTTTATAGTAAAGCAATGATTAATTTTTCATGAGGAGGCTTTAAATTTG 305
Db |||||||

Qy 1094 TGCCTTTTTCATATTTTATTCATATTCATTTAGTTTGTAACTGCTTTTAGGGAGA 1153
Db |||||||

Qy 304 TGCCTTTTTCATATTTTATTCATATTCATTTATGTTTGTAACTGCTTTTAGGGAGA 245
Db |||||||

Qy 1154 TAAATATATGTTATAAATAGTTTGGGGGAATAATTTGTCAAAGAGGATAATTAAT 1213
Db |||||||

Qy 244 TAAATATATGTTATAAATAGTTTGGGGGAATAATTTGTCAAAGAGGATAATTAAT 185
Db |||||||

Qy 1214 TACGTGCTTCTGTTATTAGATAAAGAGAGAAGCTACGCTGCATATTCAGAGTTGTA 1273
Db |||||||

Qy 184 TACGTGCTTCTGTTATTAGATAAAGAGAGAAGCTACGCTGCATATTCAGAGTTGTA 125
Db |||||||

Qy 1274 CCTTAACATTTGGTGAACATTTTCTAAGATTTTCAAAAGGATATGTGTAATTTAGA 1333
Db |||||||

Qy 124 CCTTAACATTTGGTGAACATTTTCTAAGATTTTCAAAAGGATATGTGTAATTTAGA 65
Db |||||||

Qy 1334 AATCATACCACTGTCCTTAACCTTGGTAAACAAACTGTTCTTAAATAAAGTATTTAATCAT 1393
Db |||||||

Qy 64 AATCATACCACTGTCCTTAACCTTGGTAAACAAACTGTTCTTAAATAAAGTATTTAATCAT 5
Db |||||||

Qy 1394 TTTA 1397
Db 4 TTTA 1

RESULT 9
BX441923
LOCUS

BX441923 1201 bp mRNA linear EST 15-MAY-2003

DEFINITION

BX441923 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CS0DF023YO02 5-PRIME, mRNA sequence.

ACCESSION

BX441923
VERSION
KEYWORDS
SOURCE

ORGANISM

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 1201)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope

COMMENT

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2532.r For
more information about this cluster, see
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DF023BH01QPI&cluster=2532.r>. Contact :
Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/Invitrogen> Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF023BH01QPI.

FEATURES

Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF023YO02"
/tissue_type="FETAL BRAIN"
/dev stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN

Query Match 24.1%; Score 346; DB 13; Length 1201;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 496; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 503 CATTATGCGAGCCCATTCAGTCTTGAGTTTCTCTGGACACCTTATGCTCTGAATC 562
Db |||||||

Qy 142 CATTATGCGAGCCCATTCAGTCTTGAGTTTCTCTGGACACCTTATGCTCTGAATC 201
Db |||||||

Qy 563 ATGAGCGAGGCTGATTCAATTTGGTGAATTTGGGTAGAAAGCAGTATGTTTCTGACATTA 622
Db |||||||

Qy 202 ATGAGCGAGGCTGATTCAATTTGGTGAATTTGGGTAGAAAGCAGTATGTTTCTGACATTA 261
Db |||||||

Qy 623 AGATGTAGTTATAGATAGTTTACGCTTTAAGTGTATGTTTATCTACTTTAAATAGA 682
Db |||||||

Qy 262 AGATGTAGTTATAGATAGTTTACGCTTTAAGTGTATGTTTATCTACTTTAAATAGA 321
Db |||||||

Qy 683 AATATAACCTTTAAGCTATTTCCACCTCTCTCCCGAGCTATCTCAAACTGGTGAATAT 742
Db |||||||

Qy 322 AATATAACCTTTAAGCTATTTCCACCTCTCTCCCGAGCTATCTCAAACTGGTGAATAT 381
Db |||||||

Qy 743 ATGAGAGATCTTGAAAGAGCTAAATAAATCACTTCACTGCTCCACTCCAGGTGAATCCGC 802
Db |||||||

Qy 382 ATGAGAGATCTTGAAAGAGCTAAATAAATCACTTCACTGCTCCACTCCAGGTGAATCCGC 441
Db |||||||

Qy 803 CCCTCCCACTGACCTAGTAGAATTTGTAATTAATCTTACTTCTAATTTCTGAATCA 862
Db |||||||

Qy 442 CCCTCCCACTGACCTAGTAGAATTTGTAATTAATCTTACTTCTAATTTCTGAATCA 501
Db |||||||

Qy 863 GTTGTGAATCTGTGCTTTATGTTTCAGAGTGTAAAGAACTTCAGTGAATTTTAAAA 922
Db |||||||

Qy 502 GTTGTGAATCTGTGCTTTATGTTTCAGAGTGTAAAGAACTTCAGTGAATTTTAAAA 561
Db |||||||

```

QY 923 ATCTGCTATCTGAGAGCAATGAATGAATCTTAAACAAGAGACTCATCTGTAGCTGTT 982
|
|
|
Db 562 ATCTGCTATCTGAGAGCAATGAATGAATCTTAAACAAGAGACTCATCTGTAGCTGTT 621
|
|
|
QY 983 TGCTGACTCCTATGAGCCC 1001
|
|
|
Db 622 TGCTGACTCCTATGAGCCC 640
|
|
|
RESULT 10
AA491000/c
LOCUS AA491000.1 448 bp mRNA linear EST 18-AUG-1997
DEFINITION aa5911.sl NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:824612 3',
mRNA sequence.
ACCESSION AA491000
VERSION AA491000.1 GI:2220173
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 448)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 436.
FEATURES
Location/Qualifiers
1..448
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:824612"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/clone_lib="NCI CGAP GCB1"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CEER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATGAGTGGAGCGCGCTCATTTTTTTTTTTTTTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
ORIGIN
Query Match 24.1%; Score 345; DB 9; Length 448;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1000 CCATAGGGTCTGCTTAGCATTAACAATAAGTTATAGTAAGCAATGTAT 1059
|
|
|
Db 414 CCATAGGGTCTGCTTAGCATTAACAATAAGTTATAGTAAGCAATGTAT 355
|
|
|
QY 1060 TANTTTTTTTCATGGAGGCTTAATAATTTTGCTCTTTTCATATTTATTCATAT 1119
|
|
|

```

```

Db 354 TAAATTTTTTTTGCATGAGAGGCTTTAAATTTGTGCTCTTTTTCATATTTTTCATAT 295
|
|
|
QY 1120 TCAATTTATGTTTGTAACTGCTTTTTTACGGAGATAAATTATATGTTTATAAATTTAGTTTG 1179
|
|
|
Db 294 TCAATTTATGTTTGTAACTGCTTTTTTACGGAGATAAATTATATGTTTATAAATTTAGTTTG 235
|
|
|
QY 1180 GGGGGAATAATTTGTGCAAGAGAGATAAATTAAATTACGTGCTTCGTATTACAGATAAA 1239
|
|
|
Db 234 GGGGGAATAATTTGTGCAAGAGAGATAAATTAAATTACGTGCTTCGTATTACAGATAAA 175
|
|
|
QY 1240 GAGAGAAGACTACGCTGCATATTCAAGAGCTTGCTTACCTTAACATTTGCTGTAACATTTTTC 1299
|
|
|
Db 174 GAGAGAAGACTACGCTGCATATTCAAGAGCTTGCTTACCTTAACATTTGCTGTAACATTTTTC 115
|
|
|
QY 1300 TAAAGATTTTCAAAAGGAATATGTGTAATTTGAGAATTCATAACCA 1344
|
|
|
Db 114 TAAAGATTTTCAAAAGGAATATGTGTAATTTGAGAATTCATAACCA 70
|
|
|
RESULT 11
BG484396
LOCUS BG484396
DEFINITION 602505037F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4618473 5',
mRNA sequence.
ACCESSION BG484396
VERSION BG484396.1 GI:13416675
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 785)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1376 row: j column: 10
High quality sequence start: 18
High quality sequence stop: 751.
FEATURES
Location/Qualifiers
1..785
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4618473"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_77"
/note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggccattagcc); Site 2: SfiI (ggccattagcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGCC-3' and 3' adaptor sequence:
5'-ATTTAGAGCGCGAGCGGCCACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH_MGC Library."
ORIGIN
Query Match 23.4%; Score 336; DB 12; Length 785;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 324 GTGTTAATATACAGATTTATCAGCATTTACAGCTAGGCACATATAGGAATGCAGC 383
|
|
|

```

Db 139 GTGGTAATACAGAAATTTATCAGGCAATACCAAGCTAGGCACATATAGGAATGCAGC 198
QY 384 ACTCAGAAATGGTTTCAATGTAGTAGTGTGCTTGTAGGTAGGGAGCTTATTCAGACA 443
Db 199 ACTCAGAAATGGTTTCAATGTAGTAGTGTGCTTGTAGGTAGGGAGCTTATTCAGACA 258
QY 444 TAGTAGATATTTCTCTCAATGCTGTGTCAATGTGCTGCTTGTGCTACCTGTTCCSC 503
Db 259 TAGTAGATATTTCTCTCAATGCTGTGTCAATGTGCTGCTTGTGCTACCTGTTCCSC 318
QY 504 ATTATGCGACCCCAATTCAGCTTTGAGTTTTCTCTCTGACACCTTATGCTCTGAATCA 563
Db 319 ATTATGCGACCCCAATTCAGCTTTGAGTTTTCTCTCTGACACCTTATGCTCTGAATCA 378
QY 564 TGAGCGAGGCTGATTCATTTGGTGAATTTGGGTAGAAAGCAGTAGTTTCTGCTGACATTA 623
Db 379 TGAGCGAGGCTGATTCATTTGGTGAATTTGGGTAGAAAGCAGTAGTTTCTGCTGACATTA 438
QY 624 GATGTAGGTTATAGATAGGTTTAAAGCTTTTAAAGTGTATGTTTTTATCTTTAAATAAGAA 683
Db 439 GATGTAGGTTATAGATAGGTTTAAAGCTTTTAAAGTGTATGTTTTTATCTTTAAATAAGAA 498
QY 684 ATATAACCTTTTAAAGCTTATCCACTCTCTCCCGAGCTATCTCAAACTGGTGGAAATATA 743
Db 499 ATATAACCTTTTAAAGCTTATCCACTCTCTCCCGAGCTATCTCAAACTGGTGGAAATATA 558
QY 744 TGGAGAGATCTTGAAGA 761
Db 559 TGGAGAGATCTTGAAGA 576

RESULT 12
BX452125 941 bp mRNA linear EST 22-MAY-2003
LOCUS BX452125 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CS0DF023Y002 5-PRIME, mRNA sequence.
ACCESSION BX452125
VERSION BX452125.1 GI:31034315
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 2532.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BAG018ZG08_CS01695_1&cluster=2532.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0BAG018ZG08_CS01695_1.
Location/Qualifiers
1. .941
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF023Y002"
/tissue_type="FETAL BRAIN"
/dev_stage="Fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

FEATURES
source

ORIGIN

Query Match 21.8%; Score 313; DB 13; Length 941;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1122 AATTTATGTTTCTCAATGCTGTGTCAATGTGCTGCTTGTGCTACCTGTTCCSC 1181
Db 23 AATTTATGTTTCTCAATGCTGTGTCAATGTGCTGCTTGTGCTACCTGTTCCSC 82
QY 1182 GCGAATAAATGTCCAAAGAGGATAATTTAATTTACGTGCTTCTGTTATTTCAGATAAAGA 1241
Db 83 GCGAATAAATGTCCAAAGAGGATAATTTAATTTACGTGCTTCTGTTATTTCAGATAAAGA 142
QY 1242 GAGAAGACTACGCTGCATATTCAGAGTTGTACCTTAAACAATGGTGAAACATTTTCTA 1301
Db 143 GAGAAGACTACGCTGCATATTCAGAGTTGTACCTTAAACAATGGTGAAACATTTTCTA 202
QY 1302 AGATTTTCAAAAGGAATATCTGTAAATTCAGAAATCATACCACTGCTTAACCTTGTAA 1361
Db 203 AGATTTTCAAAAGGAATATCTGTAAATTCAGAAATCATACCACTGCTTAACCTTGTAA 262
QY 1362 ACAAACTGTTCTTTAAATAAAGTATTTAATGATTTTAAAAAATTTAAAAAATTTAAAAA 1421
Db 263 ACAAACTGTTCTTTAAATAAAGTATTTAATGATTTTAAAAAATTTAAAAAATTTAAAAA 322
QY 1422 AAAAAAATTTAAAAA 1434
Db 323 AAAAAAATTTAAAAA 335

RESULT 13
AW189910/c

LOCUS AW189910 419 bp mRNA linear EST 22-NOV-1999
DEFINITION x11ba04.x1 NCI CGAP Ut4 Homo sapiens cDNA clone IMAGE:2675790 3',
similar to contains element LTR5 repetitive element ; , mRNA
sequence.
ACCESSION AW189910
VERSION AW189910.1 GI:6464390
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 419)
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco.
Location/Qualifiers
1. .419
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2675790"
/tissue_type="serous papillary carcinoma, high grade, 2
pooled tumors"
/clone_lib="NCI CGAP Ut-4"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.48 kb. Life Technologies catalog #:

FEATURES
source

```
ORIGIN
Query Match 11542-016"
Best Local Similarity 21.1%; Score 303; DB 10; Length 419;
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1099 TTTTTCATATTTTATTCATATTCATATTTATGTTTGTAACTGCTTTTGTAGGAGATAATT 1158
DB 303 TTTTTCATATTTTATTCATATTCATATTTATGTTTGTAACTGCTTTTGTAGGAGATAATT 244

QY 1159 ATATGTTATAAATTAGTTTGTGGGGGAATAATTGTCGAAAGAGATAATTAAATTTACGT 1218
DB 243 ATATGTTATAAATTAGTTTGTGGGGGAATAATTGTCGAAAGAGATAATTAAATTTACGT 184

QY 1219 GCTTCTGTTATTCAGATAAAGAGAGAGACACTACCTGCATATTCAGAGATTGTACCTTA 1278
DB 183 GCTTCTGTTATTCAGATAAAGAGAGAGACACTACCTGCATATTCAGAGATTGTACCTTA 124

QY 1279 ACATTTGGTGAACATTTTCTTAAGATTTTCAAAAGGAATATGTGTAAATTCAGAAATCA 1338
DB 123 ACATTTGGTGAACATTTTCTTAAGATTTTCAAAAGGAATATGTGTAAATTCAGAAATCA 64

QY 1339 TAACCACTGCTCTAACTGGTGAACAACTGTTCTTAATAAGTATTTAAATGATTTTAA 1398
DB 63 TAACCACTGCTCTAACTGGTGAACAACTGTTCTTAATAAGTATTTAAATGATTTTAA 4

QY 1399 AAA 1401
DB 3 AAA 1

RESULT 14
AL035942/c
LOCUS
DEFINITION 568 bp mRNA linear EST 04-SEP-2003
ACCESSION DKF2p564B1622.s1.564 (synonym: hfbr2) Homo sapiens cDNA clone
VERSION DKF2p564B1622.3, mRNA sequence.
KEYWORDS
SOURCE AL035942
ORGANISM Homo sapiens (human)
REFERENCE 1 (bases 1 to 568)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Wambutt, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
consortium of the German Genome Project.
rl sequence also available.
This clone (DKF2p564B1622) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
source
1. .568
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKF2p564B1622"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="X1-2blue"
/clone_lib="564 (synonym: hfbr2)"
/note="Vector: pMW1; Site_1: NotI; Site_2: SalI"

ORIGIN
Query Match 19.9%; Score 286; DB 9; Length 568;
Best Local Similarity 99.7%; Pred. NO. 0;
Matches 336; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1000 CCCATAAGGGTTCTGTCTAGCATTAAACAAATAAGGTTTATAGGTAAGCAATGAT 1059
DB 413 CCCATAAGGGTTCTGTCTAGCATTAAACAAATAAGGTTTATAGGTAAGCAATGAT 354

QY 1060 TAAATTTTTTTCAGTGAAGGGCTTTAAATTTGTGCTCTTTTTCATATTTTATTCATAT 1119
DB 353 TAAATTTTTTTCAGTGAAGGGCTTTAAATTTGTGCTCTTTTTCATATTTTATTCATAT 294

QY 1120 TCAATTTATGTTTGTAACTGCTTTTGTAGGAGATAATATATATATATAGTATTTG 1179
DB 293 TCAATTTATGTTTGTAACTGCTTTTGTAGGAGATAATATATATATATAGTATTTG 234

QY 1180 GGGGGAATAATTTGTCCAAAGAGGATAATTTAATTTACGTGCTTCTGTTATTCAGAAATAA 1239
DB 233 GGGGGAATAATTTGTCCAAAGAGGATAATTTAATTTACGTGCTTCTGTTATTCAGAAATAA 174

QY 1240 GAGAGAAGACTACGCTGCATATTCAGAGATTGTACCTTAAACATTTGGTGAACATTTTTC 1299
DB 173 GAGAGAAGACTACGCTGCATATTCAGAGATTGTACCTTAAACATTTGGTGAACATTTTTC 114

QY 1300 TAAAGATTTTCAAAAGGATATGTTAAATTTGAGAAAT 1336
DB 113 TAAAGATTTTCAAAAGGATATGTTAAATTTGAGAAAT 77

RESULT 15
CB132625
LOCUS
DEFINITION 557 bp mRNA linear EST 29-JAN-2003
K-EST0183064 L14Choick0 Homo sapiens cDNA clone L14Choick0-2-D11
5', mRNA sequence.
ACCESSION CB132625
VERSION CB132625.1 GI:28097898
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 557)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 2 row: D column: 11
High quality sequence stop: 557.
FEATURES
source
1. .557
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L14Choick0-2-D11"
/sex="M"
/cell_line="Choi-CK"
/lab_host="Top10F"
/clone_lib="L14Choick0"
/note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI;
Site_2: NotI; The library was contributed by the Soares
laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."

ORIGIN
```

```
Query Match      19.8%; Score 284; DB 14; Length 557;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 554; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 674 AAAATAAGAAATATAACCTTTTAAGCTATTCCACCTCTCTCCCGAGCTATCTCAAACTG 733
Db 1 AAAATAAGAAATATAACCTTTTAAGCTATTCCACCTCTCTCCCGAGCTATCTCAAACTG 60

QY 734 GTGGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCACTGCTCCACTCCAGG 793
Db 61 GTGGAATATATGGAGAGATCTTGAAGAAGTAAATAAACCCTTCACTGCTCCACTCCAGG 120

QY 794 TGAATCCGCCCACTCCCACTGACCTAGTACTAGAAATTTGTAATTTAATACTTACCTTCTATTT 853
Db 121 TGAATCCGCCCACTCCCACTGACCTAGTACTAGAAATTTGTAATTTAATACTTACCTTCTATTT 180

QY 854 CTGAAATCAGTCTGTAAGTCTGTGCTTATGTTTCAAGGTTTAAAGAACTCAGTGAATTC 913
Db 181 CTGAAATCAGTCTGTAAGTCTGTGCTTATGTTTCAAGGTTTAAAGAACTCAGTGAATTC 240

QY 914 TTTTAAAAATCTGCTATCTTGAAGACATTGAATGAATTTCTTAAACAAGAGACTCATCT 973
Db 241 TTTTAAAAATCTGCTATCTTGAAGACATTGAATGAATTTCTTAAACAAGAGACTCATCT 300

QY 974 GTAGCTGTTTGTGACTCTCTATGAGCCCAATAAGGTTTCTGTCTTAGCAATTAACAAAT 1033
Db 301 GTAGCTGTTTGTGACTCTCTATGAG -CCCAATAAGGTTTCTGTCTTAGCAATTAACAAAT 359

QY 1034 AAGGTTTATAGGTAAAGCCCAATGATTAATTTTGTGATGGAGGGCTTTAAAAATTTG 1093
Db 360 AAGGTTTATAGGTAAAGCCCAATGATTAATTTTGTGATGGAGGGCTTTAAAAATTTG 419

QY 1094 TGCTCTTTTTCATATTTTATTCATATTCATTTATGTTTGTAACTGCTTTTTAGGGAGA 1153
Db 420 TGCTCTTTTTCATATTTTATTCATATTCATTTATGTTTGTAACTGCTTTTTAGGGAGA 479

QY 1154 TAATTATATGTTTAAATTAAGTTTGGGGGGAATAATTTGCAAGAGGATAATTTAAAT 1213
Db 480 TAATTATATGTTTAAATTAAGTTTGGGGGGAATAATTTGCAAGAGGATAATTTAAAT 539

QY 1214 TAGGTGCTTCTGTTATTC 1231
Db 540 TAGGTGCTTCTGTTATTC 557
```

Search completed: July 21, 2004, 00:02:22
Job time : 3471 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 16:59:02 ; Search time 52 Seconds
(without alignments)
108.672 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 99
Sequence: 1 MAHGVLSFLWTPYALKSX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1980s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	99.0	19	6 ADA40708	Ada40708 Human sec
2	98	99.0	19	6 ABR47713	Abr47713 Human sec
3	98	99.0	20	2 AAY07853	Aay07853 Human sec
4	51	51.5	355	2 AAW07853	Aaw07853 Human sec
5	51	51.5	355	7 AAW29179	Aaw29179 Rat CC ch
6	49	49.5	2235	7 ADD45360	Add45360 Rat Prote
7	49	49.5	2235	7 ADC56729	Adc56729 Thale cre
8	48	48.5	422	2 ADC56731	Adc56731 Thale cre
9	48	48.5	422	2 AAR67688	Aar67688 S. pombe
10	47	47.5	26	6 AAR67687	Aar67687 S. pombe
11	47	47.5	148	4 ABG14555	Abg14555 Human mus
12	47	47.5	317	2 AAR48689	Aar48689 Novel hum
13	47	47.5	317	2 AAR48689	Aar48689 G-protein
14	47	47.5	456	7 AAW02661	Aaw02661 G-protein
15	47	47.5	466	1 AAD48620	Ad48620 Human pro
16	47	47.5	466	1 AAP96203	Aap96203 Human mus
17	47	47.5	466	4 AAB20483	Ab20483 Cholinerg
18	47	47.5	466	4 ABB56361	Abb56361 Non-endog
19	47	47.5	466	6 ABB99723	Abb99723 Amino aci
20	47	47.5	466	7 ABP81853	Abp81853 Human mus
21	47	47.5	466	7 ADE54491	Ad54491 Rat Prote
22	47	47.5	466	7 ADD48618	Add48618 Rat Prote
23	47	47.5	466	7 ADE54493	Ad54493 Human pro
24	47	47.5	496	5 ABG31502	Abg31502 Muscarini
25	46.5	47.0	500	5 AAB31297	Ab31297 Hydrophob
			32	4 AAM99595	Aam99595 Human exc

26	46.5	47.0	32	4 AAM42418	Aam42418 Human kid
27	46	46.5	333	2 AAR48691	Aar48691 G-protein
28	46	46.5	333	2 AAW02663	Aaw02663 G-protein
29	46	46.5	345	7 ADC86023	Adc86023 Human GPC
30	46	46.5	422	2 AAR88131	Aar88131 Aureobasi
31	46	46.5	422	2 AAR88130	Aar88130 Aureobasi
32	46	46.5	478	5 AAU97292	Aau97292 Reference
33	46	46.5	478	5 AAU97293	Aau97293 Polymorph
34	46	46.5	478	6 ABB99725	Abb99725 Amino aci
35	46	46.5	478	6 ABP81854	Abp81854 Human mus
36	46	46.5	479	1 AAP92606	Aap92606 Human mus
37	46	46.5	479	4 ABB56363	Abb56363 Non-endog
38	45	45.5	28	2 AAW40000	Aaw40000 Peptide e
39	45	45.5	209	3 AAB58414	Aab58414 Lung canc
40	45	45.5	295	4 AAG80106	Aag80106 Human CCR
41	45	45.5	355	2 AAR52749	Aar52749 C-C chemo
42	45	45.5	355	2 AAW25751	Aaw25751 Human MIP
43	45	45.5	355	2 AAW26588	Aaw26588 Human MIP
44	45	45.5	355	3 AAB20571	Aab20571 Human CC-
45	45	45.5	355	6 ABU09082	Abu09082 Human che

ALIGNMENTS

RESULT 1

ADA40708
ID ADA40708 standard; protein; 19 AA.

XX
AC ADA40708;

DT 20-NOV-2003 (first entry)

XX
DE Human secreted protein.

KW Human; secreted protein; cancer; hyperproliferative disorder;

KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;

KW anaemia; allergic reaction; asthma; cardiovascular disorder;

KW wound healing; cytostatic; immunosuppressive; neurotropic; neuroprotective;

KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;

KW vulnery; cardiant; gene therapy.

XX Homo sapiens.

OS Homo sapiens.

FN WO2002102993-A2.

XX 27-DEC-2002.

PD 19-MAR-2002; 2002WO-US008123.

XX 21-MAR-2001; 2001US-0277340P.

PR 19-JUL-2001; 2001US-0306171P.

XX 13-NOV-2001; 2001US-0331287P.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Ruben SM;

XX WPI; 2003-175238/17.

XX New human secreted proteins and nucleic acid molecules, useful for

XX preparing a diagnostic or pharmaceutical composition for diagnosing,

XX preventing or treating cancer or other hyperproliferative disorder,

XX asthma, allergies or AIDS.

PT Claim 1; SEQ ID NO 1090; 3205pp; English.

XX The invention relates to novel genes ADA39629-ADA40565 and proteins

XX ADA40566-ADA41501 for human secreted proteins, useful for preventing,

XX treating or ameliorating medical conditions e.g. by protein or gene

XX therapy. The polypeptides, nucleic acid molecules, antibodies or their

XX fragments, and agonists or antagonists that bind to the polypeptide are

XX useful for preparing a diagnostic or pharmaceutical composition for

diagnosing or treating cancer or other hyperproliferative disorder. The polypeptides and nucleic acid molecules are also useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating cancer or other hyperproliferative disorders including neoplasms, autoimmune disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic anaemia), haematopoietic or haematological disorders (e.g. anaemia, thrombocytopenia), allergic reactions including asthma or eczema, inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory bowel disease or Crohn's disease), neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's disease), cardiovascular disorders (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial, fungal or viral infections including HIV/AIDS), or wound healing and disorders of epithelial cell proliferation. The nucleic acids are also useful for chromosome identification, radiation hybrid mapping or long-range restriction mapping, as molecular weight markers, or as hybridization or diagnostic probes. The polypeptides and antibodies are useful for providing immunological probes for differential identification of the tissues immunohistochemistry assays. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Query Match 99.0%; Score 98; DB 6; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAHSLVSLFLWTPYALKS 19
Db 1 MAHSLVSLFLWTPYALKS 19

RESULT 2
ABR47713
ID ABR47713 standard; protein; 19 AA.
XX
AC ABR47713;
XX
DT 12-JUN-2003 (first entry)
XX
DE Human secreted protein, SEQ ID 604.
XX
KW Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
KW vulnary; antiinflammatory; nootropic; neuroprotective;
KW antiparkinsonian; gene therapy; human; cardiovascular disorder.
XX
OS Homo sapiens.
XX WO200295010-A2.
XX
PD 28-NOV-2002.
XX
PF 19-MAR-2002; 2002WO-US009785.
XX
FR 21-MAR-2001; 2001US-0277340P.
FR 19-JUL-2001; 2001US-0306171P.
FR 13-NOV-2001; 2001US-0331287P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2003-129429/12.
XX
PT Novel human secreted proteins, useful for detecting, preventing,
PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular
PT disorders such as arrhythmia.
XX
PS Claim 13; SEQ ID NO 604; 1801pp; English.
XX
XX The present invention relates to novel human secreted proteins (ABR47633-

ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins and their coding sequences are useful for the preparation of a diagnostic or pharmaceutical composition for diagnosing or treating a cardiovascular disorder (e.g. arrhythmia, tachycardia, cardiac arrest, coronary arteriosclerosis and myocardial ischaemia), neural disorders, immune system disorders, muscular disorders, reproductive disorders, gastrointestinal disorders, pulmonary disorders, renal disorders, CC proliferative disorders and/or cancerous diseases and conditions, for wound healing and epithelial cell proliferation, to treat inflammation or infection, for treating thrombosis and arteriosclerosis, for treating or preventing neural damage which occurs in neuronal disorders or neurodegenerative conditions such as Alzheimer's disease and Parkinson's disease, to enhance bone and periodontal regeneration and aid in tissue transplants or bone grafts, to prevent skin aging or hair loss, to stimulate growth and differentiation of haematopoietic cells and bone marrow cells when used in combination with other cytokines, to maintain organs before transplantation or for supporting cell culture of primary tissues, to increase or decrease differentiation or proliferation of embryonic stem cells, or to modulate mammalian characteristics or metabolism. Note: The sequence data for this patent was published in electronic format and is available from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 99.0%; Score 98; DB 6; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.8e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAHSLVSLFLWTPYALKS 19
Db 1 MAHSLVSLFLWTPYALKS 19

RESULT 3
AA07853
ID AA07853 standard; protein; 20 AA.
XX
AC AA07853;
XX
DT 06-JUL-1999 (first entry)
XX
DE Human secreted protein fragment encoded from gene 2.
XX
KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
KW arthritis; malignancy; digestive; endocrine; infection.
XX
OS Homo sapiens.
XX
XX WO9918208-A1.
XX
PD 15-APR-1999.
XX
PF 01-OCT-1998; 98WO-US020775.
XX
XX 02-OCT-1997; 97US-0060833P.
PR 02-OCT-1997; 97US-0060836P.
PR 02-OCT-1997; 97US-0060837P.
PR 02-OCT-1997; 97US-0060838P.
PR 02-OCT-1997; 97US-0060839P.
PR 02-OCT-1997; 97US-0060843P.
PR 02-OCT-1997; 97US-0060862P.
PR 02-OCT-1997; 97US-0060866P.
PR 02-OCT-1997; 97US-0060874P.
PR 02-OCT-1997; 97US-0060880P.
PR 02-OCT-1997; 97US-0060884P.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
 XX Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;
 PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;
 PI Lafleur DW, Shi Y;
 XX WPI; 1999-264022/22.
 DR N-PSDB; AAX37452.
 XX
 PT New isolated human genes and the secreted polypeptides they encode.
 XX
 PS Claim 1b; Page 276; 368pp; English.
 XX
 CC This invention describes novel isolated human genes and the secreted
 CC proteins they encode. The products of the invention are useful for
 CC preventing, treating or ameliorating medical conditions, e.g. by protein
 CC or gene therapy. Also pathological conditions can be diagnosed by
 CC determining the amount of the new polypeptides in a sample or by
 CC determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the 101 polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, tumours,
 CC neurodegenerative disorders, developmental abnormalities and fetal
 CC deficiencies, blood disorders, leukemias, diseases of the immune system,
 CC autoimmune diseases, hepatic and renal disease, lymphomas, inflammation,
 CC allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate
 CC disease, skeletal or cardiac muscle disorders, pulmonary disorders,
 CC transplant rejection, disorders involving osteoclasts such as
 CC osteoporosis, arthritis or malignancies, digestive/endocrine disorders,
 CC infections and AIDS. The human secreted proteins of the invention are
 CC represented in AAY07852-Y07993 and the encoding nucleic acids are
 CC represented in AAX37451-X37552
 XX
 SQ Sequence 20 AA;
 XX
 Query Match 99.0%; Score 98; DB 2; Length 20;
 Best Local Similarity 100.0%; Pred. No. 1.9e-09;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 MAHSVLSFLLWTPYALKS 19
 DB 1 MAHSVLSFLLWTPYALKS 19
 XX
 RESULT 4
 AAW29179
 ID AAW29179 standard; protein; 355 AA.
 XX
 AC AAW29179;
 XX
 DT 19-DEC-1997 (first entry)
 XX
 DE Rat CC chemokine receptor.
 XX
 KW rat; CC chemokine receptor; screen; binding; ligand.
 XX
 OS Rattus rattus.
 XX
 PN JP09227599-A.
 XX
 PD 02-SEP-1997.
 XX
 PF 22-FEB-1996; 96JP-00035192.
 XX
 PR 22-FEB-1996; 96JP-00035192.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 DR WPI; 1997-486426/45.
 DR N-PSDB; AAT86839.
 XX
 CC chemokine receptor protein - useful to screen for novel binding
 PT compounds.

XX
 PS Claim 1; Page 20-21; 26pp; Japanese.
 XX
 CC This sequence is a rat CC chemokine receptor. The receptor can be used to
 CC screen for novel binding compounds and for preparation of antibodies or
 CC antiserum
 XX
 SQ Sequence 355 AA;
 XX
 Query Match 51.5%; Score 51; DB 2; Length 355;
 Best Local Similarity 66.7%; Pred. No. 3.5;
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 XX
 QY 3 AHSVLSFLLWTPYAL 17
 DB 243 AITLFFLLWTPYNL 257
 XX
 RESULT 5
 ADD45360
 ID ADD45360 standard; protein; 355 AA.
 XX
 AC ADD45360;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 DE Rat Protein NP_065417, SEQ ID NO 10793.
 XX
 KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
 XX
 OS Rattus norvegicus.
 XX
 PN WO2003016475-A2.
 XX
 PD 27-FEB-2003.
 XX
 PF 14-AUG-2002; 2002WO-US025765.
 XX
 PR 14-AUG-2001; 2001US-0312147P.
 PR 01-NOV-2001; 2001US-0346382P.
 PR 26-NOV-2001; 2001US-0333347P.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 PA (FARB) BAYER AG.
 XX
 PI Woolf C, D'urso D, Befort K, Costigan M;
 XX
 DR WPI; 2003-268312/26.
 DR GENBANK; NP_065417.
 XX
 PT New composition comprising two or more isolated polypeptides, useful for
 PT preparing a medicament for treating pain in an animal.
 XX
 PS Claim 1; Page; 1017pp; English.
 XX
 CC The invention discloses a composition comprising two or more isolated rat
 CC or human polynucleotides or a polynucleotide which represents a fragment,
 CC derivative or allelic variation of the nucleic acid sequence. Also
 CC claimed are a vector comprising the novel polynucleotide, a host cell
 CC comprising the vector, a method for identifying a nucleotide sequence
 CC which is differentially regulated in an animal subjected to pain and a
 CC kit to perform the method, an array, a method for identifying an agent
 CC that increases or decreases the expression of the polynucleotide sequence
 CC that is differentially expressed in neuronal tissue of a first animal
 CC subjected to pain, a method for identifying a compound which regulates
 CC the expression of a polynucleotide sequence which is differentially
 CC expressed in an animal subjected to pain, a method for identifying a
 CC compound that regulates the activity of one or more of the
 CC polynucleotides, a method for producing a pharmaceutical composition, a
 CC method for identifying a compound or small molecule that regulates the
 CC activity in an animal of one or more of the polypeptides given in the
 CC specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat protein (shown in Table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 355 AA;

Query Match 51.5%; Score 51; DB 7; Length 355;
Best Local Similarity 66.7%; Pred. No. 3.5;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 AHSVLSFLLWTPYAL 17
| : : | | | | | |
Db 243 AITLFFLLWTPYNL 257

RESULT 6

ADC56729
ID ADC56729 standard; protein; 2235 AA.

XX
AC ADC56729;

XX
DT 18-DEC-2003 (first entry)

XX Thale cress cryptic precocious (CRP) protein.

XX
KW thale cress; cryptic precocious; CRP; plant; CRP-1D; vegetative growth;
KW transgenic; FT gene; Chromosome 4; leaf formation; flowering;
KW seed formation.

XX Arabidopsis thaliana.

XX JP2003024071-A.

XX 28-JAN-2003.

XX 20-JUN-2001; 2001JP-00187291.

XX 20-JUN-2001; 2001JP-00187291.

XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX WPI; 2003-516419/49.

XX N-PSDB; ADC56726, ADC56727, ADC56728.

XX Novel variant cryptic precocious-1D gene which when expressed in a plant
PT reduces vegetative growth period of the plant, useful for reducing the
PT growth period of seed crops such as rice and millet, and ornamental
PT plants.

PS Claim 9; SEQ ID NO 4; 98pp; Japanese.

XX This invention relates to a cryptic precocious (CRP) gene isolated from a
CC plant genome and a variant thereof referred to as the CRP-1D gene, and
CC which when expressed in a plant shortens the vegetative growth period of
CC the plant. Specifically, it refers to a transgenic plant that
CC overexpresses the FT gene and has an introduced CRP gene from chromosome
CC 4 of Arabidopsis thaliana or the variant CRP-1D. These transgenic plants
CC are useful for reducing the vegetative growth period of horticultural
CC plants, such that the plants start flowering in a short period of time.
CC Furthermore, they are used for commercially important seed crops such as
CC rice, millet and barley, fruit producing plants such as cucumber, cherry
CC and tomato and other crops including cocoa and coffee. The CRP-1D mutant
CC shortens the vegetative growth period, and thus avoids leaf formation
CC leading to greater flowering capability without affecting the other
CC normal functions of the plant, e.g. seed formation and growth of plant.
CC This polypeptide sequence is the thale cress CRP protein of the

CC invention.

XX
SQ Sequence 2235 AA;

Query Match 49.5%; Score 49; DB 7; Length 2235;
Best Local Similarity 41.2%; Pred. No. 55;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHSVLSFLLWTPYALKS 19

| : : | : | | | | |

Db 1529 AYSIIAFLWRPFVNS 1545

RESULT 7

ADC56731

ID ADC56731 standard; protein; 2235 AA.

XX
AC ADC56731;

XX
DT 18-DEC-2003 (first entry)

XX Thale cress SNP variant cryptic precocious protein (CRP-1D).

XX
KW thale cress; cryptic precocious; CRP; plant; CRP-1D; vegetative growth;
KW transgenic; FT gene; Chromosome 4; leaf formation; flowering;
KW seed formation; single nucleotide polymorphism.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

XX
FH Misc-difference 99

XX
FT /note= "Wild type Ser replaced by Phe due to SNP"

XX JP2003024071-A.

XX 28-JAN-2003.

XX 20-JUN-2001; 2001JP-00187291.

XX 20-JUN-2001; 2001JP-00187291.

XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX WPI; 2003-516419/49.

XX N-PSDB; ADC56730.

XX Novel variant cryptic precocious-1D gene which when expressed in a plant
PT reduces vegetative growth period of the plant, useful for reducing the
PT growth period of seed crops such as rice and millet, and ornamental
PT plants.

XX Example; SEQ ID NO 6; 98pp; Japanese.

XX This invention relates to a cryptic precocious (CRP) gene isolated from a
CC plant genome and a variant thereof referred to as the CRP-1D gene, and
CC which when expressed in a plant shortens the vegetative growth period of
CC the plant. Specifically, it refers to a transgenic plant that
CC overexpresses the FT gene and has an introduced CRP gene from chromosome
CC 4 of Arabidopsis thaliana or the variant CRP-1D. These transgenic plants
CC are useful for reducing the vegetative growth period of horticultural
CC plants, such that the plants start flowering in a short period of time.
CC Furthermore, they are used for commercially important seed crops such as
CC rice, millet and barley, fruit producing plants such as cucumber, cherry
CC and tomato and other crops including cocoa and coffee. The CRP-1D mutant
CC shortens the vegetative growth period, and thus avoids leaf formation
CC leading to greater flowering capability without affecting the other
CC normal functions of the plant, e.g. seed formation and growth of plant.
CC This polypeptide sequence is the thale cress variant CRP-1D protein
CC containing a S99F mutation caused by a single nucleotide polymorphism
CC given in ADC56730 of the invention.

XX
SQ Sequence 2235 AA;

Query Match 49.5%; Score 49; DB 7; Length 2235;
Best Local Similarity 41.2%; Pred. No. 55;
Matches 7; Conservative 6; Mismatches 4; Indels

QY 3 AHSVLSFLLWTPYALKS 19
 ||:::||::||:
Db 1529 AYSIIAFVLWRPFVSNS 1545

RESULT 8
AAR67688
ID AAR67688 standard; protein; 422 AA.

AC	AAR67688;	16-OCT-2003	(revised)
XX		04-AUG-1995	(first entry)
DT			
DT			

DE *S. pombe* spaur1S gene product.

KW Aureobasidin-resistance; aureobasidin-sensitivity; spaur1s gene; mycosis;
KW antimycotic; fungicide; diagnosis; therapy.

OS Schizosaccharomyces pombe; JY 745.

PN AU9463129-A.
XX
PD 01-DEC-1994.

16-MAY-1994; 94AU-00063129.
PF XX
24-MAY-1993; 93JP-00142523.
PR
28-DEC-1993; 93JP-00348893.
PR

PA (TAKI) TAKARA SHUZO CO LTD.

PI XX DR DR
Okada T, Takesako K, Kato I;
WPI: 1995-023286/04.
N-PSDE; AAO75953.

New genes regulating aureobasidin sensitivity - used to develop prods. for the diagnosis and treatment of diseases such as mycoses.

PS Disclosure; Page 64-66; 110pp; English.

Schizosaccharomyces pombe JY 745⁺, which is sensitive to aureobasidin, was mutagenized with EMS and a genomic library of a resistant strain was prepared. A DNA fragment containing a resistant gene, spaurLr (given in AAQV592), was isolated. Using this gene as a probe, DNA containing a sensitive gene, spaurIs (AAQV5953), was obtained from the sensitive strain. A G1053T mutation leading to a G240C conversion gave rise to the resistance. [Updated on 16-OCT-2003 to standardise OS field]

Sequence 422 AA;

```
Query Match      48.5%; Score 48; DB 2; Length 422;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 5; Indels
```

QY 4 HSVLSFLLWTPYAL 17
||:| | | | | :
Db 142 HSILDILAWVPYGV 155

RESULT 9
AAR67687
ID AAR67687 standard; protein: 422 AA.

AC AAR67687;

	16-OCT-2003 (revised)	04-AUG-1995 (first entry)
DT		
DT		

XX
DE
XX
KW
KW
XX
OS

PN AU9463129-A.
XX
PD 01-DEC-1994.

16-MAY-1994; 94AU-00063129.

PR 24-MAY-1993; 93JP-00142523.
PR 28-DEC-1993; 93JP-00348893.

PA (TAKI) TAKARA SHUZO CO LTD.

PI Okada T, Takesato K, Kato I;
XX
DR WPI; 1995-023286/04.
DR N-PSDB: AAO75952.

PT New genes regulating aureobasidin sensitivity - used to develop prods.
PT for the diagnosis and treatment of diseases such as mycoses.

PS Disclosure; Page 60-62; 110pp; English.

CC Schizosaccharomyces pombe JY 745, which is sensitive to aureobasidin, was
CC mutagenized with EMS and a genomic library of a resistant strain was
CC prepared. A DNA fragment containing a resistant gene, spaurin (given in
CC AAQ75952), was isolated. Using this gene as a probe, DNA containing a
CC sensitive gene, spauris (AAQ75953), was obtained from the sensitive
CC strain. A G1053T mutation leading to a G240C conversion gave rise to the
CC resistance. (Updated on 16-OCT-2003 to standardise OS field)

Sequence 422 AA;

Query Match 48.5%; Score 48; DB 2; Length 422;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 7; Conservative 2; Mismatches 5; Indels

```

QY      4 HSVLSFLLWTPYAL 17
         ||:| | | | | :
Db     142 HSILDILAWVPYGV 155

```

RESULT 10
ABP55266
ID ABP55266 standard; peptide: 26 AA.

AC ABP55266:

28-JAN-2003 (first entry)

DE Human muscarinic acetylcholine m2 receptor TM 4 peptide SEQ ID NO:19
XX

Human; muscarinic acetylcholine m2 receptor; histamine H2 receptor;
 KW serotonin 5HT2A receptor; dopamine D2 receptor; rhodopsin; GPCR;
 KW adrenergic b2 receptor; transmembrane; G protein-coupled receptor;
 KW protein co-ordinate data; three dimensional structure.

OS Homo sapiens.

PN WO200279784-A1.

PD 10-OCT-2002

PF 01-APR-2002; 2002WO-JP003264.

PR 30-MAR-2001; 2001JP-00101510.

PA (SUNR) SUNTORY LTD.
 XX Ishiguro M;
 XX WPI; 2003-058447/05.
 DR Construction of three-dimensional structural models of G protein-coupled
 XX receptor and binding ligand complex as well as activation intermediates.
 PT applicable in identifying, examining, searching, evaluating and designing
 PT drugs.
 XX
 XX Example 2; Fig 18; 390pp; Japanese.
 PS
 XX The present invention describes a method for constructing 3D (three-
 CC dimensional) structural models of the activation intermediate of G
 CC protein-coupled receptors (GPCRs) for use in identifying, examining,
 CC searching, evaluating and designing agonists or antagonists as ligand
 CC binding to the GPCRs. The constructed structural models are applicable in
 CC identifying, examining, searching, evaluating and designing drugs,
 CC particularly of orphan receptors in terms of full and partial agonists,
 CC antagonists by de novo design, as well as for use in the design of
 CC constitutively active receptors by screening ligands with mutants to
 CC develop and provide highly effective agents with no side-reactions. The
 CC present sequence represents a human GPCR transmembrane domain peptide
 CC which is used in an example from the present invention
 XX
 XX Sequence 26 AA;
 SQ
 Query Match 47.5%; Score 47; DB 6; Length 26;
 Best Local Similarity 58.8%; Pred. No. 0.97;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MARHSLVSLFLWTPYAL 17
 :||| ||||:|
 Db 10 IAAAWLSFLWAPAIL 26
 RESULT 11
 ABG14555
 ID ABG14555 standard; protein; 148 AA.
 XX AC ABG14555;
 XX 18-FEB-2002 (first entry)
 DT
 XX Novel human diagnostic protein #14546.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 KW
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 FN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US008631.
 PF
 XX 31-MAR-2000; 2000US-00540217.
 PR
 XX 23-AUG-2000; 2000US-00649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR
 XX N-PSDB; AAS78742.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX

PS Claim 20; SEQ ID NO 44914; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 148 AA;
 SQ
 Query Match 47.5%; Score 47; DB 4; Length 148;
 Best Local Similarity 64.7%; Pred. No. 6.4;
 Matches 11; Conservative 1; Mismatches 3; Indels 2; Gaps 1;
 QY 2 AAHSVLSFLWTPYALK 18
 :||| |||||
 Db 50 ASHSVSSFLAW--YQLK 64
 RESULT 12
 AAR48689
 ID AAR48689 standard; protein; 317 AA.
 XX AC AAR48689;
 XX 05-JUN-1996 (first entry)
 DT
 XX G-protein coupled human m2 muscarinic acetylcholine receptor protein.
 DE G-protein coupled receptor; ligand binding assay; transmembrane domain;
 KW psychotic disorder; schizophrenia; dopamine; cAMP; adenosine; thrombin;
 KW muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;
 KW rhodopsin; opsin; odorant; cytomegalovirus.
 XX
 XX Homo sapiens.
 OS
 XX WO9405695-A1.
 FN
 XX 17-MAR-1994.
 PD
 XX 09-SEP-1993; 93WO-US008528.
 PF
 XX 10-SEP-1992; 92US-00943236.
 PR
 XX (UYNV) UNIV NEW YORK STATE.
 PA
 XX Murphy RB, Schuster DI;
 PI
 XX WPI; 1994-101120/12.
 DR
 XX Polypeptides of G-coupled receptor proteins (GPRs) - useful for binding
 PT GPR ligands or modulating GPR binding.
 XX
 XX Disclosure; Page 61; 160pp; English.
 PS
 XX Proteins AAR48685-R48758 represent a range of G-protein coupled receptor
 CC proteins selected from cAMP, adenosine, muscarinic acetylcholine,
 CC

CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
 CC odorant, cytomegaloviral and other G-protein coupled receptors. The
 CC receptor proteins were used to design polypeptides, pref. based on the
 CC transmembrane domains, for use in G-protein coupled receptor ligand
 CC binding assays. The polypeptide fragments retain biological activity such
 CC as binding a GPR ligand or modulating GPR ligand binding to a GPR (see
 CC AAR48759-R48758, AAR50569-R50807 and AAR89189-R89195 for examples of
 CC polypeptide fragments). The polypeptide fragments can be used in
 CC compositions for treating subjects suffering from a pathology related to
 CC a GPR abnormality e.g. a psychotic disorder such as schizophrenia
 XX
 SQ Sequence 317 AA;

Query Match 47.5%; Score 47; DB 2; Length 317;
 Best Local Similarity 58.8%; Pred. No. 15;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSHVLSFLWTPVAL 17
 :|||:|||||
 Db 111 IAAWVLSFILWAPAIL 127

RESULT 13

AAW02661
 ID AAW02661 standard; peptide; 317 AA.
 AC AAW02661;

XX
 XX

DT 25-MAR-2003 (revised)

DT 12-NOV-1996 (first entry)

XX
 XX

DE G-protein coupled human m2 muscarinic acetylcholine receptor.

XX
 XX

KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
 KW schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;
 KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
 KW odorant; cytomegalovirus; serotonergic.

XX
 OS Homo sapiens.

XX
 XX US5508384-A.

PN
 XX 16-APR-1996.

PD
 XX 09-SEP-1993; 93US-00118270.

PF
 XX 10-SEP-1992; 92US-00943236.

PR
 XX (UUNY) UNIV NEW YORK STATE.

XX
 PA Schuster DI, Murphy RB;

XX
 PI WPI; 1996-208785/21.

XX
 DR New dopamine receptor peptide - useful as antipsychotic agent, e.g. for
 PT treating schizophrenia.

XX
 PT Disclosure; Col 49-52; 184pp; English.

XX
 PS

XX
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

CC
 CC

Query Match 47.5%; Score 47; DB 2; Length 317;
 Best Local Similarity 58.8%; Pred. No. 15;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSHVLSFLWTPVAL 17
 :|||:|||||
 Db 111 IAAWVLSFILWAPAIL 127

RESULT 14

ADD48620
 ID ADD48620 standard; protein; 456 AA.

XX
 AC ADD48620;

XX
 DT 29-JAN-2004 (first entry)

XX
 DE Human Protein AAK68113, SEQ ID NO 14326.

XX
 KW Human; pain; neuronal tissue; gene therapy;

XX
 KW spinal segmental nerve injury; chronic constriction injury; CCI;

XX
 KW spared nerve injury; SNI; Chung.

XX
 OS Homo sapiens.

XX
 PN WO2003016475-A2.

XX
 PD 27-FEB-2003.

XX
 PF 14-AUG-2002; 2002WO-US025765.

XX
 PR 14-AUG-2001; 2001US-0312147P.

XX
 PR 01-NOV-2001; 2001US-0346382P.

XX
 PR 26-NOV-2001; 2001US-0333347P.

XX
 PA (GEHO) GEN HOSPITAL CORP.

XX
 PA (FARB) BAYER AG.

XX
 PI Woolf C, D'urso D, Befort K, Costigan M;

XX
 XX WPI; 2003-268312/26.

XX
 DR GENEANK; AAK68113.

XX
 DR

XX
 PT

XX
 PT

XX
 XX

XX
 PS

XX
 XX

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

XX
 CC

The invention discloses a composition comprising two or more isolated rat or human polynucleotides or a polynucleotide which represents a fragment, derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (Chung), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note:

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 456 AA;

Query Match 47.5%; Score 47; DB 7; Length 456;
Best Local Similarity 58.8%; Pred. No. 22;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAHSVLSFLWTPYAL 17
:|||:|||||:
Db 144 IAAAWLSFILWAPAIL 160

RESULT 15
AAP96203
ID AAP96203 standard; protein; 466 AA.
XX
AC AAP96203;
XX
XX 25-MAR-2003 (revised)
DT 17-DEC-2001 (revised)
DT 27-SEP-1989 (first entry)
XX
XX Human muscarinic acetylcholine m2 receptor.
DE
XX Muscarinic acetylcholine receptor; m2; drug screening; diagnosis;
KW drug design.
XX
XX Homo sapiens.
OS
XX USN7241971-N.
PN
XX 14-MAR-1989.
PD
XX 08-SEP-1988; 88US-00241971.
XX
XX 08-SEP-1988; 88US-00241971.
PR
XX (USSH) NAT INST OF HEALTH.
XX PA
XX Bonner T;
XX PI
XX WPI; 1989-165452/22.
DR N-FSDB; AAN92069.
DR
XX Cloned genes for muscarinic acetyl:choline receptors - for drug screening
PT and diagnostic use.
PT
XX Disclosure; Page; 21pp; English.
PS
XX Cell lines producing this MAR is useful in screening for drugs with
CC improved specificity for the part. MAR type. The amino acid sequence may
CC be useful for rational drug design and antibodies to the sequences may be
CC useful for diagnostic purposes. See also AAN92068-N92067 and AAN90086.
CC (Note: Revised entry submitted to correct the patent number format of US
CC Government-owned NTIS applications to prevent clashes with ongoing US
CC granted patent numbers. For further information please visit the Derwent
CC web site at www.derwent.com/dwpi/updates/ntis_us.html.) (Updated on 25-
CC MAR-2003 to correct PI field.)
XX
SQ Sequence 466 AA;

Query Match 47.5%; Score 47; DB 1; Length 466;
Best Local Similarity 58.8%; Pred. No. 22;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAAHSVLSFLWTPYAL 17
:|||:|||||:
Db 144 IAAAWLSFILWAPAIL 160

Search completed: July 19, 2004, 17:05:25
Job time : 61 secs

RESULT 3
US-08-492-459-4
; Sequence 4, Application US/08492459
; Patent No. 6015689
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: REGULATION OF AUROBASIDIN SENSITIVITY IN FUNGUS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/492,459
; FILING DATE: June 20, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:

Query Match	48.5%;	Score 48;	DB 3;	Length 423;
Best Local Similarity	50.0%;	Pred. No. 7.3;		
Matches	7;	Conservative	2;	Mismatches 5; Indels
QY	4	HSVLSFLLWTPYAL	17	
				:
Db	142	HSILDIILAWPYGV	155	

```

RESULT 5
US-08-423-752-4
; Sequence 4, Application US/08423752
; Patent No. 6022949
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; TITLE OF INVENTION: AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/423,752
; FILING DATE: April 18, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
;
US-08-423-752-4

Query Match 48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels

QY 4 HSVLSFLWTPYAL 17
Db 142 HGIIDILAWPYGV 155

RESULT 6
US-08-716-873-7
; Sequence 7, Application US/08716873
; Patent No. 6194166
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

```

```

; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-716-873-16

Query Match          48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy      4 HSVLSFLWTPYAL 17
Db      142 HSILDILAWPYGV 155

RESULT 8
US-08-716-873-18
; Sequence 18, Application US/08716873
; Patent No. 6194166
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/716.873
; FILING DATE: September 20, 1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-716-873-18

Query Match          48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy      4 HSVLSFLWTPYAL 17
Db      142 HSILDILAWPYGV 155

RESULT 9
US-09-368-431-7
; Sequence 7, Application US/09368431
; Patent No. 6294651
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/368,431
; FILING DATE: August 5, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/716,873
; FILING DATE: September 20, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-368-431-7

Query Match          48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy      4 HSVLSFLWTPYAL 17
Db      142 HSILDILAWPYGV 155

RESULT 10
US-09-368-431-16
; Sequence 16, Application US/09368431
; Patent No. 6294651
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/368,431
; FILING DATE: August 5, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/716,873
; FILING DATE: September 20, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-368-431-16

Query Match          48.5%; Score 48; DB 3; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy      4 HSVLSFLWTPYAL 17
Db      142 HSILDILAWPYGV 155
```

; FILING DATE: August 5, 1999
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/716,873
 ; FILING DATE: September 20, 1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-721-8200
 ; TELEFAX: 202-721-8250
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 422 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-09-368-431-16

Query Match 48.5%; Score 48; DB 3; Length 422;
 Best Local Similarity 50.0%; Pred. No. 7.3;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
 ||:| | | | |
 Db 142 HSILDILAWPYGV 155

RESULT 11
 US-09-368-431-18
 ; Sequence 18, Application US/09368431
 ; Patent No. 6294651
 ; GENERAL INFORMATION:
 ; APPLICANT: Takashi OKADO et al.
 ; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
 ; TITLE OF INVENTION: (AS AMENDED)
 ; NUMBER OF SEQUENCES: 50
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
 ; STREET: 2033 K Street, N.W., #800
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/368,431
 ; FILING DATE: August 5, 1999
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/716,873
 ; FILING DATE: September 20, 1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-721-8200
 ; TELEFAX: 202-721-8250
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 18:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 422 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear

; MOLECULE TYPE: peptide
 ; US-09-368-431-18

Query Match 48.5%; Score 48; DB 3; Length 422;
 Best Local Similarity 50.0%; Pred. No. 7.3;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
 ||:| | | | |
 Db 142 HSILDILAWPYGV 155

RESULT 12
 US-09-414-006-2
 ; Sequence 2, Application US/09414006
 ; Patent No. 6348577
 ; GENERAL INFORMATION:
 ; APPLICANT: Takashi OKADO et al.
 ; TITLE OF INVENTION: REGULATION OF AUREOBASIDIN SENSITIVITY (AS
 ; TITLE OF INVENTION: AMENDED)
 ; NUMBER OF SEQUENCES: 36
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
 ; STREET: 2033 K Street, N.W., #800
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20006
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Wordperfect 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/414,006
 ; FILING DATE: October 7, 1999
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/492,459
 ; FILING DATE: June 20, 1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/243,403
 ; FILING DATE: May 16, 1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Warren M. Cheek, Jr.
 ; REGISTRATION NUMBER: 33,367
 ; REFERENCE/DOCKET NUMBER:
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-721-8200
 ; TELEFAX: 202-721-8250
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 422
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-09-414-006-2

Query Match 48.5%; Score 48; DB 4; Length 422;
 Best Local Similarity 50.0%; Pred. No. 7.3;
 Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
 ||:| | | | |
 Db 142 HSILDILAWPYGV 155

RESULT 13
 US-09-414-006-4
 ; Sequence 4, Application US/09414006
 ; Patent No. 6348577

```
;
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: REGULATION OF AUROBASIDIN SENSITIVITY (AS
; TITLE OF INVENTION: AMENDED)
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/414,006
; FILING DATE: October 7, 1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/492,459
; FILING DATE: June 20, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-414-006-4

Query Match 48.5%; Score 48; DB 4; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSLSFLLWTPYAL 17
Db 142 HSILDILAWPYGV 155

;
; RESULT 14
; US-09-447-223-2
; Sequence 2, Application US/09447223
; Patent No. 6432664
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/447,223
; FILING DATE: 23-No. 6432664-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: <Unknown>
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-414-006-4
```

```
;
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/447,223
; FILING DATE: 23-No. 6432664-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: <Unknown>
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
; US-09-447-223-2

Query Match 48.5%; Score 48; DB 4; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSLSFLLWTPYAL 17
Db 142 HSILDILAWPYGV 155

;
; RESULT 15
; US-09-447-223-4
; Sequence 4, Application US/09447223
; Patent No. 6432664
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; AUROBASIDIN SENSITIVITY
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/447,223
; FILING DATE: 23-No. 6432664-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX: <Unknown>
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-414-006-4
```

;
; LENGTH: 422
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-447-223-4

Query Match 48.5%; Score 48; DB 4; Length 422;
Best Local Similarity 50.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSLFLLWTPYAL 17
||| | | | | |
Db 142 HSILDILAWPYGV 155

Search completed: July 19, 2004, 17:07:25
Job time : 20 secs

This Page Blank (2021)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:06:36 ; Search time 43 Seconds
(without alignments)
145.377 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 99

Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US05_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	99.0	20	14	US-10-195-730-116
2	51	51.5	355	14	US-10-376-564-13
3	49	49.5	364	16	US-10-437-963-163690
4	48	48.5	422	10	US-09-951-217-7
5	48	48.5	422	10	US-09-951-217-16
6	48	48.5	422	10	US-09-951-217-18
7	47	47.5	196	14	US-10-017-161-794
8	47	47.5	466	11	US-09-826-509-515
9	47	47.5	466	14	US-10-225-567A-190
10	47	47.5	488	13	US-10-029-009-7
11	47	47.5	496	13	US-10-029-009-34
12	47	47.5	500	13	US-10-029-009-19
13	46.5	47.0	32	15	US-10-242-355-332
14	46	46.5	345	14	US-10-017-161-542
15	46	46.5	345	15	US-10-292-798-476

16 46 46.5 478 14 US-10-225-567A-193
17 46 46.5 479 11 US-09-826-509-519
18 46 46.5 490 10 US-09-992-238-7
19 46 46.5 1907 14 US-10-156-761-10430
20 45.5 46.0 681 15 US-10-369-493-13394
21 45 45.5 154 16 US-10-437-963-119491
22 45 45.5 209 9 US-09-925-302-752
23 45 45.5 209 12 US-09-925-302-752
24 45 45.5 279 12 US-10-246-583-58
25 45 45.5 279 15 US-10-174-364-58
26 45 45.5 355 9 US-09-961-068-1
27 45 45.5 355 9 US-09-960-547-1
28 45 45.5 355 12 US-03-893-512-11
29 45 45.5 355 12 US-10-039-659-13
30 45 45.5 355 14 US-10-225-567A-62
31 45 45.5 355 14 US-10-245-850-2
32 45 45.5 355 14 US-10-164-649-53
33 45 45.5 355 14 US-10-376-564-14
34 45 45.5 355 14 US-10-239-423-62
35 45 45.5 355 14 US-10-439-845-9
36 45 45.5 355 15 US-10-452-015-2
37 45 45.5 375 14 US-10-219-834-78
38 45 45.5 500 16 US-10-437-963-154068
39 44 44.4 771 12 US-10-425-114-60213
40 44 44.4 824 12 US-10-282-122A-52599
41 43 43.4 68 12 US-10-424-599-176608
42 43 43.4 102 12 US-10-424-599-172079
43 43 43.4 141 12 US-10-424-599-188984
44 43 43.4 246 15 US-10-369-493-12318
45 43 43.4 284 9 US-09-867-550-1872

ALIGNMENTS

RESULT 1
US-10-195-730-116
; Sequence 116, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 116
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (20)
; OTHER INFORMATION: Xaa equals stop translation
US-10-195-730-116

Query Match 99.0%; Score 98; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.7e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAHSVLSFLWTPYALKS 19
| | | | | | | | | | | | | | | | | | | | | |
Db 1 MAHSVLSFLWTPYALKS 19

RESULT 2

US-10-376-564-13

; Sequence 13, Application US/10376564

; Publication No. US20030180302A1

; GENERAL INFORMATION:

; APPLICANT: Wolf, Eckhard

; APPLICANT: Werner, Sabine

; APPLICANT: Halle, Jörn-Peter

; APPLICANT: Regenbogen, Johannes

; APPLICANT: Goppelt, Andreas

; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for

; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound

; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically

; TITLE OF INVENTION: Active Substances

; FILE REFERENCE: 50125/014003

; CURRENT APPLICATION NUMBER: US/10/376,564

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 09/886,319

; PRIOR FILING DATE: 2001-06-20

; PRIOR APPLICATION NUMBER: US 60/222,081

; PRIOR FILING DATE: 2000-08-01

; PRIOR APPLICATION NUMBER: DE 100 30 149.5-41

; PRIOR FILING DATE: 2000-06-20

; NUMBER OF SEQ ID NOS: 86

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13

; LENGTH: 355

; TYPE: PRT

; ORGANISM: Mus musculus

US-10-376-564-13

Query Match 51.5%; Score 51; DB 14; Length 355;

Best Local Similarity 66.7%; Pred. No. 13;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AHSVLSFLWTPYAL 17

Db 243 AITLLFLLWTPYNL 257

RESULT 3

US-10-437-963-163690

; Sequence 163690, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 163690

; LENGTH: 364

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(364)

; OTHER INFORMATION: unsure at all Xaa locations

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_62662C.1.pep

US-10-437-963-163690

Query Match 49.5%; Score 49; DB 16; Length 364;

Best Local Similarity 62.5%; Pred. No. 27;

Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MAHSVLSFLWTPYA 16

Db 148 MAANDGLSLLWMPYS 163

RESULT 4

US-09-951-217-7

; Sequence 7, Application US/09951217

; Publication No. US20030050440A1

; GENERAL INFORMATION:

; APPLICANT: Takashi OKADO et al.

; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY

; (AS AMENDED)

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: WENDROTH, LIND & PONACK, L.L.P.

; STREET: 2033 K Street, N.W., #800

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20006

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/951,217

; FILING DATE: 06-Sep-2001

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 09/368,431

; FILING DATE: August 5, 1999

; APPLICATION NUMBER: 08/716,873

; FILING DATE: September 20, 1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Warren M. Cheek, Jr.

; REGISTRATION NUMBER: 33,367

; REFERENCE/DOCKET NUMBER: <Unknown>

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-721-8200

; TELEFAX: 202-721-8250

; TELEX: <Unknown>

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 422 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-09-951-217-7

Query Match 48.5%; Score 48; DB 10; Length 422;

Best Local Similarity 50.0%; Pred. No. 44;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 4 HSVLSFLWTPYAL 17

Db 142 HSLDILAWPYGV 155

RESULT 5

US-09-951-217-16

; Sequence 16, Application US/09951217

; Publication No. US20030050440A1

; GENERAL INFORMATION:

; APPLICANT: Takashi OKADO et al.

; TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY

; (AS AMENDED)

; NUMBER OF SEQUENCES: 50

; CORRESPONDENCE ADDRESS:

ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/951,217
FILING DATE: 06-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/368,431
FILING DATE: August 5, 1999
APPLICATION NUMBER: 08/716,873
FILING DATE: September 20, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-951-217-16
Query Match 48.5%; Score 48; DB 10; Length 422;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 4 HSVLSFLWTPYAL 17
Db 142 HSILDILAWPYGV 155
RESULT 6
US-09-951-217-18
Sequence 18, Application US/09951217
Publication No. US2003005040A1
GENERAL INFORMATION:
APPLICANT: Takashi OKADO et al.
TITLE OF INVENTION: GENE REGULATING AUROBASIDIN SENSITIVITY
(AS AMENDED)
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/951,217
FILING DATE: 06-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/368,431
FILING DATE: August 5, 1999
APPLICATION NUMBER: 08/716,873
FILING DATE: September 20, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER: <Unknown>
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-09-951-217-18
Query Match 48.5%; Score 48; DB 10; Length 422;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 4 HSVLSFLWTPYAL 17
Db 142 HSILDILAWPYGV 155
RESULT 7
US-10-017-161-794
Sequence 794, Application US/10017161
Publication No. US2003014368A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 084335/0152
CURRENT APPLICATION NUMBER: US/10/017,161
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 794
LENGTH: 196
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: MOD_RES
LOCATION: (25)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (52)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (59)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (67)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (73)
OTHER INFORMATION: Variable amino acid
FEATURE:

; NAME/KEY: MOD RES
; LOCATION: (78)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (101)..(104)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (126)..(127)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (134)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (147)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (166)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (172)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (179)
; OTHER INFORMATION: Variable amino acid
; US-10-017-161-794

Query Match 47.5%; Score 47; DB 14; Length 196;
Best Local Similarity 58.8%; Pred. No. 30;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
: || ||||| :
Db 3 IAAWVLSFILWAPAIL 19

RESULT 8
US-09-826-509-515
; Sequence 515, Application US/09826509
; Publication No. US20030204073A1
; GENERAL INFORMATION:
; APPLICANT: Lehmann-Bruinsma, Karin
; APPLICANT: Liaw, Chen W.
; APPLICANT: Lin, I-Lin
; TITLE OF INVENTION: No. US20030204073A1-Endogenous, Constitutively Activated Known G
; TITLE OF INVENTION: Protein-Coupled Receptors
; FILE REFERENCE: AREN-207
; CURRENT APPLICATION NUMBER: US/09/826,509
; CURRENT FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/195,747
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 09/170,496
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 589
; SOFTWARE: PatentIn Version 2.1
; SEQ ID NO 515
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-826-509-515

Query Match 47.5%; Score 47; DB 11; Length 466;
Best Local Similarity 58.8%; Pred. No. 68;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
: || ||||| :
Db 3 IAAWVLSFILWAPAIL 19

Db 144 IAAWVLSFILWAPAIL 160
RESULT 9
US-10-225-567A-190
; Sequence 190, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: LifeSpan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burmer, Glenn C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTOR
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 190
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-190

Query Match 47.5%; Score 47; DB 14; Length 466;
Best Local Similarity 58.8%; Pred. No. 68;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
: || ||||| :
Db 144 IAAWVLSFILWAPAIL 160

RESULT 10
US-10-029-009-7
; Sequence 7, Application US/10029009
; Publication No. US20020164617A1
; GENERAL INFORMATION:
; APPLICANT: Felsch, Jason S.
; APPLICANT: Annis, David Allen
; APPLICANT: Kalghatgi, Krishna
; APPLICANT: Nash, Huw M.
; TITLE OF INVENTION: Affinity Selection-Based Screening of Hydrophobic Proteins
; FILE REFERENCE: 111025.173 US2
; CURRENT APPLICATION NUMBER: US/10/029,009
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/258,970
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Human
US-10-029-009-7

Query Match 47.5%; Score 47; DB 13; Length 488;
Best Local Similarity 58.8%; Pred. No. 71;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
: || ||||| :
Db 166 IAAWVLSFILWAPAIL 182

RESULT 11
US-10-029-009-34
; Sequence 34, Application US/10029009
; Publication No. US20020164617A1
; GENERAL INFORMATION:
; APPLICANT: Felsch, Jason S.

```

; APPLICANT: Annis, David Allen
; APPLICANT: Kalghatgi, Krishna
; APPLICANT: Nash, Huw M.
; TITLE OF INVENTION: Affinity Selection-Based Screening of Hydrophobic Proteins
; FILE REFERENCE: 111025.173 US2
; CURRENT APPLICATION NUMBER: US/10/029,009
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/258,970
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Met-Melitin SS-FLAG M1 tag-m2 mChR sequence
US-10-029-009-34

```

```

Query Match          47.5%; Score 47; DB 13; Length 496;
Best Local Similarity 58.8%; Pred. No. 72;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1 MAHSVLSFLWTPYAL 17
      :|||:|||||
Db      174 IAAAWLSFILWAPAIL 190

```

```

RESULT 12
US-10-029-009-19
; Sequence 19, Application US/10029009
; Publication No. US20020164617A1
; GENERAL INFORMATION:
; APPLICANT: Pelsch, Jason S.
; APPLICANT: Annis, David Allen
; APPLICANT: Kalghatgi, Krishna
; APPLICANT: Nash, Huw M.
; TITLE OF INVENTION: Affinity Selection-Based Screening of Hydrophobic Proteins
; CURRENT APPLICATION NUMBER: US/10/029,009
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/258,970
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Human
US-10-029-009-19

```

```

Query Match          47.5%; Score 47; DB 13; Length 500;
Best Local Similarity 58.8%; Pred. No. 73;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

```

```

QY      1 MAHSVLSFLWTPYAL 17
      :|||:|||||
Db      184 IAAAWLSFILWAPAIL 200

```

```

RESULT 13
US-10-242-355-332
; Sequence 332, Application US/10242355
; Publication No. US20030235831A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC003C1
; CURRENT APPLICATION NUMBER: US/10/242,355
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,897
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065

```

```

; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 332
; LENGTH: 32
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (26)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-242-355-332

```

```

Query Match          47.0%; Score 46.5; DB 15; Length 32;
Best Local Similarity 56.2%; Pred. No. 6.3;
Matches 9; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

```

```

QY      1 MAHSVLSF-LLWTPY 15
      :|||:|||||
Db      17 MAPHLSFXLCWLPY 32

```

```

RESULT 14
US-10-017-161-542
; Sequence 542, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 542
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (4)..(37)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (49)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (271)
; OTHER INFORMATION: Variable amino acid
; FEATURE:

```

NAME/KEY: MOD RES
LOCATION: (276)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (285)..(318)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (330)
OTHER INFORMATION: Variable amino acid
US-10-017-161-542

Query Match 46.5% Score 46; DB 14; Length 345;
Best Local Similarity 58.8%; Pred. No. 72;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAASHSVLSFLWTPYAL 17
Db 150 IAAAVLSFVLWAPAIL 166

RESULT 15
US-10-292-798-476
Sequence 476, Application US/10292798
Publication No. US20030235833A1
GENERAL INFORMATION:
APPLICANT: SUWA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
FILE REFERENCE: 084335/166
CURRENT APPLICATION NUMBER: US/10/292,798
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 10/017,161
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: JP 2001-246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2070
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 476
LENGTH: 345
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MOD RES
LOCATION: (4)..(37)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (49)..(49)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (271)..(271)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (276)..(276)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (285)..(318)
OTHER INFORMATION: Variable amino acid
FEATURE:
NAME/KEY: MOD RES
LOCATION: (330)..(330)
OTHER INFORMATION: Variable amino acid
US-10-292-798-476

Query Match 46.5% Score 46; DB 15; Length 345;
Best Local Similarity 58.8%; Pred. No. 72;

Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Qy 1 MAASHSVLSFLWTPYAL 17
Db 150 IAAAVLSFVLWAPAIL 166
Search completed: July 19, 2004, 17:12:04
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:03:23 ; Search time 16 Seconds
(without alignments)
120.239 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 99

Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:*

1: Pir1:*

2: Pir2:*

3: Pir3:*

4: Pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	51.5	355	2 149339	macrophage inflamm
2	49	49.5	2124	201526	hypothetical prote
3	48	48.5	422	2 T38764	aureobasidin a res
4	47.5	48.0	355	2 I51319	RH2 opsin - green
5	47	47.5	466	2 A40972	muscarinic acetyl
6	47	47.5	466	2 JH0197	muscarinic acetyl
7	47	47.5	466	2 S10126	muscarinic acetyl
8	47	47.5	466	2 S10836	muscarinic acetyl
9	47	47.5	466	2 A27386	muscarinic acetyl
10	46	46.5	478	2 C29514	muscarinic acetyl
11	46	46.5	479	2 S10127	muscarinic acetyl
12	46	46.5	479	2 S33776	muscarinic acetyl
13	46	46.5	490	2 A35546	muscarinic acetyl
14	45.5	46.0	355	1 A46191	iodopsin homolog -
15	45	45.5	355	2 A45177	chemokine (C-C) re
16	44.5	44.9	355	2 A42347	opsin, green-sensi
17	44	44.4	301	2 T33218	hypothetical prote
18	44	44.4	308	2 T36751	hypothetical prote
19	44	44.4	557	2 S21733	PAC3 protein - hum
20	44	44.4	563	2 T14242	NADH2 dehydrogenas
21	44	44.4	649	2 B38129	bo-type ubiquinol
22	43	43.4	354	2 I58186	probable G protein
23	43	43.4	414	2 AF3423	cyanate transport
24	43	43.4	459	2 T11149	NADH2 dehydrogenas
25	43	43.4	870	2 T30110	hypothetical prote
26	42.5	42.9	349	2 B45229	opsin, green-sensi
27	42	42.4	281	2 S54137	probable movement
28	42	42.4	305	2 T45942	hypothetical prote
29	42	42.4	305	2 A84799	hypothetical prote

30 42 42.4 355 2 JC4304 orphan G protein-c
31 42 42.4 359 2 149341 MIP-1 alpha recept
32 42 42.4 360 2 JC2443 chemokine (C-C) re
33 42 42.4 374 2 I38450 chemokine (C-C) re
34 42 42.4 589 2 AH0370 probable binding p
35 42 42.4 729 2 A91038 hypothetical prote
36 42 42.4 742 2 D85882 hypothetical prote
37 42 42.4 768 2 H65013 yfex protein - Esc
38 42 42.4 772 2 H85054 hypothetical prote
39 41 41.4 278 2 B81601 ABC transporter, p
40 41 41.4 278 2 F72064 solute binding pro
41 41 41.4 278 2 A86558 solute binding pro
42 41 41.4 310 2 G86490 probable bionin ho
43 41 41.4 323 2 B83215 conserved hypothet
44 41 41.4 376 2 S28853 opsin - red swamp
45 41 41.4 378 2 S28789 opsin Rh4 - fruit

ALIGNMENTS

RESULT 1

149339
macrophage inflammatory protein-1 alpha receptor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: I49339
R:Gao, J.L.; Murphy, P.M.
J. Biol. Chem. 270, 17494-17501, 1995
A:Title: Cloning and differential tissue-specific expression of three mouse beta chemoki-
A:Reference number: I49339; MUID:95340546; PMID:7542241
A:Accession: I49339
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-355 <RES>
A:Cross-references: EMBL:U28404; NID:9881547; PIDN:AAA89153.1; PID:9881548
C:Superfamily: vertebrate rhodopsin

Query Match 51.5%; Score 51; DB 2; Length 355;
Best Local Similarity 66.7%; Pred. No. 1.3;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 AHSVLSFLWTPYAL 17
|:::| | | | | | | |
Db 243 AITLFFELWTPYNL 257

RESULT 2

T01526
hypothetical protein A_IC005I10.24 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Nov-1999
C:Accession: T01526
R:Andrews, S.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of A. thaliana IC005I10.
A:Reference number: Z14347
A:Accession: T01526
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2124 <AND>
A:Cross-references: EMBL:AF013293; NID:92252823; PID:92252843
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 11/3; 49/3; 84/3; 141/2; 256/2; 989/3; 1019/2; 1889/2; 1951/2; 1973/3
A:Note: A_IC005I10.24
C:Superfamily: Arabidopsis thaliana hypothetical protein A_IC005I10.24

Query Match 49.5%; Score 49; DB 2; Length 2124;
Best Local Similarity 41.2%; Pred. No. 16;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 3 AHSVLSFLLWTPYALKS 19
|:|:|:|:|:|:|:
Db 1438 AYSIIAFVLWRPFVNS 1454

RESULT 3
T38764
autobasidin a resistance protein - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C/Accession: T38764
R/Gentiles, S.; Churche, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, January 1996
A/Reference number: Z21810
A/Accession: T38764
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-422 <GEN>
A/Cross-references: EMBL:Z69086; PIN:CAA93163.1; GSPDB:GN00066; SPDB:SPAC3H8.06
A/Experimental source: strain 972h; cosmid c3H8
C/Genetics:
A/Gene: aurl; SPDB:SPAC3H8.06
A/Map position: 1

Query Match	48.5%	Score 48;	DB 2;	Length 422;
Best Local Similarity	50.0%;	Pred. No. 4.7;		
Matches	7;	Conservative	2;	Mismatches
				5;
				Indels
				0;
				Gaps
				0;

Qy 4 HSVLSFLLWTPYAL 17
||:| | | | | :
Db 142 HSILDILAWVPYGV 155

RESULT 4
I51319
RH2 opsin - green anole
C/Species: Anolis carolinensis (green anole)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999
C/Accession: I51319
R/Kawamura, S.; Yokoyama, S.
J. Mol. Evol. 40, 594-600, 1995
A/Rittle: Paralogous origin of the rhodopsinlike opsin genes in lizards.
A/Reference number: I51319; MUID:95371134; PMID:7643409
A/Accession: I51319
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-355 <RAW>
A/Cross-references: GB:S79167; NID:g1042072; PIDN:AAB35062.1; PID:g1042073
C/Genetics:
A/Gene: rh2Ac

Query Match	48.0%	Score 47.5	DB 2	Length 355
Best Local Similarity	76.9%	Pred. No. 4.8		
Matches 10	Conservative	1	Mismatches	1
			Indels	1
			Gaps	1

```
QY      6 VLSFLL-WTPYAL 17
        |||||
Db      258 VLGFLLAWTPYAM 270

RESULT 5
A40972
muscarinic acetylcholine receptor M2 - chicken
C:Species: Gallus gallus (Chicken)
C>Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 24-Nov-1999
C:Accession: A40972
R:Tietje, K.M.; Nathanson, N.M.
J. Biol. Chem. 266, 17382-17387, 1991
```

A>Title: Embryonic chick heart expresses multiple muscarinic acetylcholine receptor subtypes with high affinity for pirenzepine.
A:Reference number: A40972; MUID: 91373358; PMID: 1840593
A:Accession: A40972
A:Molecule type: DNA
A:Residues: 1-466 <TIE>
A:Cross-references: GB:M73217
C:Genetics:
A:Gene: cm2
A:Introns: #status absent
C:Superfamily: vertebrate rhodopsin
C:Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; phospholipase C
F:26-51/Domain: transmembrane #status predicted <TM1>
F:64-88/Domain: transmembrane #status predicted <TM2>
F:101-122/Domain: transmembrane #status predicted <TM3>
F:142-165/Domain: transmembrane #status predicted <TM4>
F:187-210/Domain: transmembrane #status predicted <TM5>
F:389-409/Domain: transmembrane #status predicted <TM6>
F:421-442/Domain: transmembrane #status predicted <TM7>

Query Match 47.5%; Score 47; DB 2; Length 466;
Best Local Similarity 58.8%; Pred. No. 7.6;
Matches 10: Conservative 2; Mismatches 5: Indels

```
Qy      1 MAAHSVLSFLLWTPYAL 17
        :|||:|||||
Db     147 IAAAVLSFLLWAPAIL 163
```

```

RESULT 6
JH0197
muscarinic acetylcholine receptor M2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 24-Nov-1999
C:Accession: JH01971, D37121
R:Iai, J.; Bloom, J.W.; Yamamura, H. I.; Roeske, W.R.
Life Sci. 47, 1001-1013, 1990
A:Title: Amplification of the rat m2 muscarinic receptor gene by the polymerase chain reaction
A:Reference number: JH0197; MUID:91041524; PMID:2172674
A:Accession: JH0197
A:Molecule type: DNA
A:Residues: 1-466 <LAI>
R:Kartenbach, E.; Curtis, C.A.M.; Pedder, E.K.; Aitken, A.; Harris, A.C.M.;
J. Biol. Chem. 265, 13702-13708, 1990
A:Title: Muscarinic acetylcholine receptors. Peptide sequencing identifies
A:Reference number: A37121; MUID:90337982; PMID:2380182

```

Query Match 47.5%; Score 47; DB 2; Length 466;
Best Local Similarity 58.8%; Pred. No. 7.6;
Matches 10; Conservative 2; Mismatches 5; Indels

```

QY      1  MAHSVLSFLLWTPYAL 17
        :|||:||:|
DB     144  IAAVWLSFILWAPAIL 160

```

RESULT 7
S10126
muscarinic acetylcholine receptor M2 - human
C1Species: Homo sapiens (man)

C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 24-Nov-1999
A;Accession: S10126
R;Peralta, E.G.; Ashkenazi, A.; Winslow, J.W.; Smith, D.H.; Ramachandran, J.; Capon, D.J.
EMBO J. 6, 3923-3929, 1987
A;Title: Distinct primary structures, ligand-binding properties and tissue-specific expression of the human interleukin-3 receptor
A;Reference number: S04326; MUID:88166632; PMID:3443095
A;Accession: S10126
A;Molecule type: DNA
A;Residues: 1-466 <PER>
A;Cross-references: EMBL:X15264; NID:g32319; PIDN:CAA33335.1; PID:g32320
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; phospholipase C
F;23-48/Domain: transmembrane #status predicted <TM1>
F;60-89/Domain: transmembrane #status predicted <TM2>
F;98-119/Domain: transmembrane #status predicted <TM3>
F;139-162/Domain: transmembrane #status predicted <TM4>
F;184-207/Domain: transmembrane #status predicted <TM5>
F;389-409/Domain: transmembrane #status predicted <TM6>
F;421-442/Domain: transmembrane #status predicted <TM7>
F;2,6,9/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 47.5%; Score 47; DB 2; Length 466;
Best Local Similarity 58.8%; Pred. No. 7.6;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAHSLVSLFLWTPYAL 17
DB 144 TAAAWLSFILWAPAIL 160
RESULT 8
S10856
muscarinic acetylcholine receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 24-Nov-1999
C;Accession: S10856
R;Gocayne, J.; Robinson, D.A.; FitzGerald, M.G.; Chung, F.Z.; Kerlavage, A.R.; Lentes, K.
Proc. Natl. Acad. Sci. U.S.A. 84, 8296-8300, 1987
A;Title: Primary structure of rat cardiac beta-adrenergic and muscarinic cholinergic receptors
A;Reference number: S10855; MUID:88068581; PMID:2825184
A;Accession: S10856
A;Molecule type: mRNA
A;Residues: 1-466 <GOC>
A;Cross-references: EMBL:J03025; NID:g203461; PIDN:AAA40926.1; PID:g203462
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; phospholipase C
F;23-48/Domain: transmembrane #status predicted <TM1>
F;61-85/Domain: transmembrane #status predicted <TM2>
F;98-119/Domain: transmembrane #status predicted <TM3>
F;139-162/Domain: transmembrane #status predicted <TM4>
F;184-207/Domain: transmembrane #status predicted <TM5>
F;389-409/Domain: transmembrane #status predicted <TM6>
F;421-442/Domain: transmembrane #status predicted <TM7>
Query Match 47.5%; Score 47; DB 2; Length 466;
Best Local Similarity 58.8%; Pred. No. 7.6;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAHSLVSLFLWTPYAL 17
DB 144 TAAAWLSFILWAPAIL 160
RESULT 9
A27386
muscarinic acetylcholine receptor, cardiac - pig
N;Alternate names: muscarinic acetylcholine receptor M2
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 15-Dec-1988 #sequence_revision 15-Dec-1988 #text_change 31-Dec-2000
A;Accession: A27386; A25656
R;Peralta, E.G.; Winslow, J.W.; Peterson, G.L.; Smith, D.H.; Ashkenazi, A.; Ramachandran, J.
Science 236, 600-605, 1987
A;Title: Primary structure and biochemical properties of an M-2 muscarinic receptor.

A;Reference number: A27386; MUID:87206169; PMID:3107123
A;Accession: A27386
A;Molecule type: DNA; mRNA
A;Residues: 1-466 <PER>
A;Cross-references: GB:M16331; NID:g164311; PIDN:AAA30986.1; PID:g164313
A;Experimental source: atrial muscle
A;Note: the protein sequence derived from the mRNA clones differs from that of the genomic DNA
R;Castro, J.
FEBS Lett. 209, 367-372, 1986
A;Title: Primary structure of porcine cardiac muscarinic acetylcholine receptor deduced from complementary DNA
A;Reference number: A25656; MUID:87080790; PMID:3792556
A;Accession: A25656
A;Molecule type: mRNA
A;Residues: 1-329, 'K', 331-466 <KUB>
A;Cross-references: GB:X04708; NID:g1859; PIDN:CAA28413.1; PID:g1860
A;Experimental source: cardiac muscle
C;Superfamily: vertebrate rhodopsin
C;Keywords: cardiac muscle; G protein-coupled receptor; glycoprotein; heart; neurotransmitter
F;23-48/Domain: transmembrane #status predicted <TM1>
F;61-85/Domain: transmembrane #status predicted <TM2>
F;98-119/Domain: transmembrane #status predicted <TM3>
F;139-162/Domain: transmembrane #status predicted <TM4>
F;184-207/Domain: transmembrane #status predicted <TM5>
F;389-409/Domain: transmembrane #status predicted <TM6>
F;421-442/Domain: transmembrane #status predicted <TM7>
Query Match 47.5%; Score 47; DB 2; Length 466;
Best Local Similarity 58.8%; Pred. No. 7.6;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAHSLVSLFLWTPYAL 17
DB 144 TAAAWLSFILWAPAIL 160
RESULT 10
C29514
muscarinic acetylcholine receptor M4 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 24-Nov-1999
C;Accession: C94518; C94293; E37121; C29514
R;Bonner, T.I.
submitted to GenBank, July 1987
A;Reference number: A94518
A;Accession: C94518
A;Molecule type: mRNA
A;Residues: 1-478 <BO1>
R;Bonner, T.I.; Buckley, N.J.; Young, A.C.; Brann, M.R.
Science 237, 527-532, 1987
A;Title: Identification of a family of muscarinic acetylcholine receptor genes.
A;Reference number: A94293; MUID:87263421; PMID:3037705
A;Accession: C94293
A;Molecule type: mRNA
A;Residues: 1-233;373-478 <BO2>
A;Experimental source: cerebral cortex
A;Note: only a part of the protein translation is given; none of the nucleotide sequence
R;Kuttenbach, E.; Curtis, C.A.M.; Pedder, E.K.; Aitken, A.; Harris, A.C.M.; Hulme, E.C.
J. Biol. Chem. 265, 13702-13708, 1990
A;Title: Muscarinic acetylcholine receptors. Peptide sequencing identifies residues involved in binding of the antagonist atropine
A;Reference number: A37121; MUID:90337982; PMID:2380182
A;Accession: E37121
A;Status: preliminary
A;Molecule type: protein
A;Residues: 68-130 <KUR>
C;Superfamily: vertebrate rhodopsin
C;Keywords: G protein-coupled receptor; glycoprotein; neurotransmitter receptor; phospholipase C
F;31-56/Domain: transmembrane #status predicted <TM1>
F;69-93/Domain: transmembrane #status predicted <TM2>
F;106-127/Domain: transmembrane #status predicted <TM3>
F;147-170/Domain: transmembrane #status predicted <TM4>
F;192-215/Domain: transmembrane #status predicted <TM5>
F;401-421/Domain: transmembrane #status predicted <TM6>
F;433-454/Domain: transmembrane #status predicted <TM7>

F;8,13/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 46.5%; Score 46; DB 2; Length 478;
Best Local Similarity 58.8%; Pred.No. 11;
Matches 10; Conservative 2; Mismatches 5; Indels

```

Qy      1  MAAHSVLSFLLWTPYAL 17
        :|| ||||:|||||
Db     152  IAAAWVLSFVLWAPAIL 168

```

RESULT 11

muscarinic acetylcholine receptor M3 - human
C/Species: Homo sapiens (man)
C/Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 24-Nov-1999
C/Accession: S10127
R/Peralta, E.G.; Ashkenazi, A.; Winslow, J.W.; Smith, D.H.; Ramachandran, J.; Capon, D.J.
EMBO J. 6, 3923-3929, 1987
A/Title: Distinct primary structures, ligand-binding properties and tissue-specific expression of the human muscarinic acetylcholine receptor subtypes M1, M2, M3 and M4
A/Reference number: S04326; MUID:68166632; PMID:3443095
A/Accession: S10127
A/Molecule type: DNA
A/Residues: 1-479 <PER>

Query Match 46.5%; Score 46; DB 2; Length 479;
Best Local Similarity 58.8%; Pred. No. 11;
Matches 10; Conservative 2; Mismatches 5; Indels

Qy 1 MAHSVLSFLLWTPYAL 17
: || || || : || ||
Db 153 IAAAWLSFVLWAPAIL 169

RESULT 12

S33776
muscarinic acetylcholine receptor m4 - mouse
C|Species: Mus musculus (house mouse)
C|Date: 08-Dec-1993 #sequence_revision 01-Dec-1995 #text_change 24-Nov-1999
C|Accession: S33776; S33135
R|van Koppen, C.J.; Lenz, W.; Nathanson, N.M.
Biochim. Biophys. Acta 1173, 342-344, 1993
A|Title: Isolation, sequence and functional expression of the mouse m4 muscarinic acetylcholine receptor cDNA
A|Reference number: S33776; MUID: S3305731; PMID: 7916637
A|Accession: S33776
A|Molecule type: DNA
A|Residues: 1-479 <KOP>
A|Cross-references: EMBL:X63473; NID:g296913; PIDN:CAA45071.1; PID:g296914

Query Match 46.5%; Score 46; DB 2; Length 479;
Best Local Similarity 58.8%; Pred. No. 11;
Matches 10; Conservative 2; Mismatches 5; Indels

```
QY      1 MAHSVLSFLLWTPYAL 17
        :|||:||:|
Db     152 IAAVWLSFVLWAPAIL 168
```

RESULT 13

A35546
muscarinic acetylcholine receptor M4 - chicken
C/Species: Gallus gallus (chicken)
C/Date: 22-Jan-1993 #sequence revision 22-Jan-1993 #text_change 21-Jul-2000
C/Accession: A35546
J/Tietje, K.M.; Goldman, P.S.; Nathanson, N.M.
J. Biol. Chem. 265, 2828-2834, 1990
A/Title: Cloning and functional analysis of a gene encoding a novel muscarinic acetylcholine receptor
A/Reference number: A35546; MUID:90153912; PMID:2154460

Query Match	46.5%	Score 46;	DB 2;	Length 490;
Best Local Similarity	52.3%;	Pred. No. 11;		
Matches 9;	Conservative	3;	Mismatches	5;
			Indels	0;
			Gaps	0;

QY 1 MAHSVLSFLLWTPYAL 17
: : : : :
pb 163 IAAWILSFLLWAPAIL 179

RESULT 14

A46191
iodopsin homolog - tokay
C:Species: Gekko gekko (tokay)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A46191
R:Kojima, D.; Okano, T.; Fukuda, Y.; Shichida, Y.; Yoshizawa, T.; Ebrey, T.G.
Proc. Natl. Acad. Sci. U.S.A. 89, 6841-6845, 1992
A:Title: Cone visual pigments are present in gekko rod cells.
A:Reference number: A46191: MUID:92157726: PMID:1379723

Query Match	46.0%	Score 45.5;	DB 1;	Length 355;
Best Local Similarity	83.3%;	Pred. No. 10;		
Matches 10;	Conservative	0;	Mismatches	1;
			Indels	1;
			Gaps	1;

Qy 6 VLSFLL-WTPYA 16
Db 258 VLGFLLAWTPYA 269

RESULT 15

A45177
chemokine (C-C) receptor 1 - human
N;Alternate names: C-C CKR-1; macrophage inflammatory protein-1-alpha receptor

C:Species: Homo sapiens (man)
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 13-Aug-1999
 C:Accession: A45177; I55671
 R:Neote, K.; DiGregorio, D.; Mak, J.Y.; Horuk, R.; Schall, T.J.
 Cell 72, 415-425, 1993
 A:Title: Molecular cloning, functional expression, and signaling characteristics of a C-
 cell 72, 415-425, 1993
 A:Reference number: A45177; MUID:93161416; PMID:7679328
 A:Accession: A45177
 A>Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-355 <NEO>
 A:Cross-references: GB:L10918; NID:g292416; PIDN:AAA36543.1; PID:g292417
 A:Experimental source: HL60 Cells
 A:Note: sequence extracted from NCBI backbone (NCBIP:124876)
 R:Gao, J.
 J. Exp. Med. 177, 1421-1427, 1993
 A:Title: Structure and functional expression of the human macrophage inflammatory 1 alpha
 A:Reference number: I55671; MUID:93240122; PMID:7683036
 A:Accession: I55671
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-355 <RES>
 A:Cross-references: GB:L10918; NID:g292416; PIDN:AAA36543.1; PID:g292417
 C:Genetics:
 A:Gene: GDB:CMKBR1; CMKR-1
 A:Cross-references: GDB:138446; OMIM:601159
 A:Map position: 3p21-3p21
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: disulfide bond; G protein-coupled receptor; glycoprotein; phosphoprotein; tr
 F;36-60/Domain: transmembrane #status predicted <TM1>
 F;71-91/Domain: transmembrane #status predicted <TM2>
 F;108-129/Domain: transmembrane #status predicted <TM3>
 F;147-171/Domain: transmembrane #status predicted <TM4>
 F;205-223/Domain: transmembrane #status predicted <TM5>
 F;240-284/Domain: transmembrane #status predicted <TM6>
 F;288-305/Domain: transmembrane #status predicted <TM7>
 F;5/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;24-273,106-183/disulfide bonds: #status predicted
 F;345/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted

Query Match 45.5%; Score 45; DB 2; Length 355;
 Best Local Similarity 58.3%; Pred. No. 12;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 6 VLSFLWTPYAL 17
 Db 246 IFFFLWTPYNL 257

Search completed: July 19, 2004, 17:06:56
 Job time : 17 secs

This Page Blank (uspto)


```

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
-!- FUNCTION: RECEPTOR FOR A C-C TYPE CHEMOKINE. BINDS TO MIP-1-ALPHA,
RANTES, AND LESS EFFICIENTLY, TO MIP-1-BETA OR MCP-1 AND
SUBSEQUENTLY TRANSDUCES A SIGNAL BY INCREASING THE INTRACELLULAR
CALCIUM IONS LEVEL. RESPONSIBLE FOR AFFECTING STEM CELL
PROLIFERATION.
-!- SUBCELLULAR LOCATION: Integral membrane protein.
-!- TISSUE SPECIFICITY: Detected in the heart, spleen, lung,
peritoneal exudate cells and leukocytes.
-!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U29678; AAA86119.1; -.
EMBL; U28404; AAA89153.1; -.
ENBL; BC011092; AAH11092.1; -.
PIR; I49339; I49339; Ccrl.
MGD; MGI:104618; Ccrl.
GO; GO:0016493; F:C-C chemokine receptor activity; IDA.
GO; GO:0005155; P:protein binding; IPI.
GO; GO:00030595; P:immune cell chemotaxis; IDA.
GO; GO:0006954; P:inflammatory response; IMP.
GO; GO:0030099; P:myeloid blood cell differentiation; IMP.
InterPro; IPR000276; GPCR_Rhodopsin.
PRINTS; PR00237; GPCRHOPOPSN.
PROSITE; PS00237; G PROTEIN RECP FL 1; 1.
PROSITE; PS0262; G PROTEIN RECP FL 2; 1.
G-protein coupled receptor; Transmembrane.
DOMAIN 1 34 EXTRACELLULAR (POTENTIAL).
TRANSMEM 35 60 1 (POTENTIAL).
DOMAIN 61 64 CYTOPLASMIC (POTENTIAL).
TRANSMEM 65 91 2 (POTENTIAL).
DOMAIN 92 107 EXTRACELLULAR (POTENTIAL).
TRANSMEM 108 129 3 (POTENTIAL).
DOMAIN 130 146 CYTOPLASMIC (POTENTIAL).
TRANSMEM 147 171 4 (POTENTIAL).
DOMAIN 172 197 EXTRACELLULAR (POTENTIAL).
TRANSMEM 198 223 5 (POTENTIAL).
DOMAIN 224 239 CYTOPLASMIC (POTENTIAL).
TRANSMEM 240 264 6 (POTENTIAL).
EXTRACELLULAR (POTENTIAL).
TRANSMEM 265 281 7 (POTENTIAL).
DOMAIN 282 305 CYTOPLASMIC (POTENTIAL).
TRANSMEM 306 355 BY SIMILARITY.
FT DISULFID 106 183
FT CONFLICT 55 55 M -> V (IN REF. 2 AND 3).
FT CONFLICT 149 149 L -> F (IN REF. 3).
FT CONFLICT 278 278 H -> Q (IN REF. 3).
SQ SEQUENCE 355 AA; 40901 MW; FCE9FF70E6F38B1 CRC64;

Query Match 51.5%; Score 51; DB 1; Length 355;
Best Local Similarity 66.7%; Pred. No. 1-2;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 AHSVLSFLWTPYAL 17
DB 243 AITLFFLLWTPYNL 257

RESULT 2
AURI_SCHPO STANDARD; PRT; 422 AA.
ID AURI_SCHPO
AC Q10142; P79014;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aurbasidin A resistance protein homolog.

```

```

GN AURI OR SPAC3H8.06.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=THR01;
RA Hashida-Okado T.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Spouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weijtens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst A., Thode G.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
"the genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-890(2002).
-!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
-!- SIMILARITY: BELONGS TO THE AURI FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB000821; BAA19190.1; -.
EMBL; Z69086; CAA93163.1; -.
PIR; T38764; T38764.
GeneDB SPombe; SPAC3H8.06; -.
InterPro; IPR008934; AcPase VanPerase.
InterPro; IPR000326; PA_FTPase.
PFam; PF01569; PAP2; 1.
SMART; SM00014; acidPPC; 1.
Transmembrane.
KW TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 136 156 POTENTIAL.
FT TRANSMEM 161 181 POTENTIAL.
FT TRANSMEM 184 204 POTENTIAL.
FT TRANSMEM 246 266 POTENTIAL.
FT TRANSMEM 293 313 POTENTIAL.
FT CONFLICT 240 240 G -> C (IN REF. 1).
SQ SEQUENCE 422 AA; 47444 MW; C24EC4E3423E442E CRC64;

Query Match 48.5%; Score 48; DB 1; Length 422;
Best Local Similarity 50.0%; Pred. No. 3.9;

```

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 HSVLSFLWTPYAL 17
 ||:| | | | |
 142 HSLDILAWPYGV 155

Db

RESULT 3

OPSB_ANOCA STANDARD; PRT; 355 AA.

ID OPSB_ANOCA

AC PS1471;

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DE Blue-sensitive opsin (Blue photoreceptor pigment) (RH2 opsin).

OS Anolis carolinensis (Green anole) (American chameleon).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Lepidosauria; Squamata; Iguania; Iguanidae; Polychrotinae; Anolis.

OX NCBI_TaxID=28377;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95371134; PubMed=7643409;

RA Kawamura S., Yokoyama S.;

RT "Paralogous origin of the rhodopsinlike opsin genes in lizards.";

RL J. Mol. Evol. 40:594-600 (1995).

CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 IMMEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 LINKED TO CIS-RETINAL. THIS OPSIN USES A VITAMIN-A2 CHROMOPHORE.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- PTM: Some or all of the carboxyl-terminal Ser or Thr residues may
 be phosphorylated (By similarity).

CC -1- MISCELLANEOUS: THIS OPSIN HAS PROBABLY AN ABSORPTION MAXIMUM AT
 503 NM.

CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 Opsin subfamily.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC

EMBL; S79167; AAB35062.1; JOINED.

EMBL; S79124; AAB35062.1; JOINED.

EMBL; S79134; AAB35062.1; JOINED.

EMBL; S79165; AAB35062.1; JOINED.

EMBL; S79166; AAB35062.1; JOINED.

PIR; I51319; I51319.

DR HSP; P02699; 1BQJ.

DR InterPro; IPR000276; GPCR_Rhodpsn.

DR InterPro; IPR001760; Opsin.

DR Pfam; PF00001; 7tm.1; 1.

DR PRINTS; PR00237; GPCRHHODOPSN.

DR PROSITE; PS00237; G PROTEIN RECF F1_1; 1.

DR PROSITE; PS50262; G-PROTEIN_RECEP_F1_2; 1.

DR PROSITE; PS00238; OPSIN; 1.

KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 Phosphorylation; G-protein coupled receptor; Lipoprotein; Palmitate.

DOMAIN 1 36
 EXTRACELLULAR.

FT TRANSMEM 37 61
 DOMAIN 62 73
 CYTOPLASMIC.

FT TRANSMEM 74 98
 DOMAIN 99 113
 EXTRACELLULAR.

FT TRANSMEM 114 133
 DOMAIN 134 152
 CYTOPLASMIC.

FT TRANSMEM 153 176
 DOMAIN 177 202
 EXTRACELLULAR.

FT TRANSMEM 203 230
 DOMAIN 231 252
 CYTOPLASMIC.

FT TRANSMEM 253 276
 DOMAIN 277 284
 EXTRACELLULAR.

FT TRANSMEM 285 309
 DOMAIN 310 355
 CYTOPLASMIC.

FT CARBOHYD 2 2
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 15 15
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 200 200
 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT DISULFID 110 187
 BY SIMILARITY.

FT BINDING 296 296
 RETINAL CHROMOPHORE (BY SIMILARITY).

FT LIPID 322 322
 S-palmitoyl cysteine (By similarity).

FT LIPID 323 323
 S-palmitoyl cysteine (By similarity).

SQ SEQUENCE 355 AA; 39983 MW; 555B671A4DE96EA6 CRC64;

Query Match 48.0%; Score 47.5; DB 1; Length 355;
 Best Local Similarity 76.9%; Pred. No. 4;
 Matches 10; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 6 VLSFLWTPYAL 17
 || | | | | | | | |
 258 VLGFLLAWTPYAM 270

Db

RESULT 4

OPSB_HEMSA STANDARD; PRT; 377 AA.

ID OPSB_HEMSA

AC Q25158;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE Compound eye opsin BCRH2.

OS Hemigrapsus sanguineus (Crab).

OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;

OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;

OC Subbrachyura; Grapsoidae; Varunidae; Hemigrapsus.

OX NCBI_TaxID=40176;

RN [1]

RP SEQUENCE FROM N.A.

RX TISSUE=Eye;

RA Sakamoto K., Hisatomi O., Tokunaga F., Eguchi E.;

RT "Two opsins from the compound eye of the crab Hemigrapsus
 sanguineus.";

RL J. Exp. Biol. 199:441-450 (1996).

CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 IMMEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 LINKED TO CIS-RETINAL. THIS OPSIN PRODUCES VISUAL PIGMENTS WITH
 MAXIMAL ABSORPTION IN THE BLUE-GREEN REGION OF THE SPECTRUM.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL OF THE SEVEN RETINULAR CELLS
 (R1-R7) FORMING THE MAIN RHABDOM IN EACH OMATIDIUM.

CC -1- PTM: Some or all of the carboxyl-terminal Ser or Thr residues may
 be phosphorylated (By similarity).

CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 Opsin subfamily.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC

EMBL; D50584; BAA09133.1; --

DR HSP; P02699; 1F88.

DR InterPro; IPR000276; GPCR_Rhodpsn.

DR InterPro; IPR001760; Opsin.

DR Pfam; PF00001; 7tm.1; 1.

DR PRINTS; PR00237; GPCRHHODOPSN.

DR PROSITE; PS00237; G PROTEIN RECF F1_1; 1.

DR PROSITE; PS50262; G-PROTEIN_RECEP_F1_2; 1.

DR PROSITE; PS00238; OPSIN; 1.

KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 Phosphorylation; G-protein coupled receptor.

DOMAIN 1 53
 EXTRACELLULAR.

FT TRANSMEM 54 78
 DOMAIN 79 1
 (POTENTIAL).

FT DOMAIN 79 90 CYTOPLASMIC.
 FT TRANSMEM 91 115 2 (POTENTIAL).
 FT DOMAIN 116 131 EXTRACELLULAR.
 FT TRANSMEM 132 151 3 (POTENTIAL).
 FT DOMAIN 152 170 CYTOPLASMIC.
 FT TRANSMEM 171 194 4 (POTENTIAL).
 FT DOMAIN 195 218 EXTRACELLULAR.
 FT TRANSMEM 219 246 5 (POTENTIAL).
 FT DOMAIN 247 281 CYTOPLASMIC.
 FT TRANSMEM 282 305 6 (POTENTIAL).
 FT DOMAIN 306 313 EXTRACELLULAR.
 FT TRANSMEM 314 338 7 (POTENTIAL).
 FT DOMAIN 339 377 CYTOPLASMIC.
 FT TRANSMEM 378 402 8 (POTENTIAL).
 FT BINDING 128 205 BY SIMILARITY.
 FT BINDING 325 325 RETINAL CHROMOPHORE (BY SIMILARITY).
 FT CARBOHYD 3 3 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 377 AA; 42114 MW; FD6CC2E0E199A256 CRC64;

Query Match 47.5%; Score 47; DB 1; Length 377;
 Best Local Similarity 69.2%; Pred. No. 5.1;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 7 LSFLLWTPYALKS 19
 ||:|||||
 Db 289 LWFICWTPYALIS 301

RESULT 5
 AC2M2_PANTR STANDARD; PRT; 440 AA.
 AC Q9N2A7;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Muscarinic acetylcholine receptor M2 (Fragment).
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate 220;
 RA Kitano T., Kobayakawa H., Saitou N.;
 RT "Silver project";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
 CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
 CC BREAKDOWN OF PHOSPHOLIPIDS & MODULATION OF POTASSIUM CHANNELS
 CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 CC ADENYLATE CYCLASE INHIBITION.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB041392; BAA94477.1;
 CC InterPro; IPR000276; GPCR_Rhodpsn.
 CC Pfam; PF00001; 7tm.1; 1.
 CC PRINTS; PR00237; GPCR_Rhodopsin
 CC PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.
 CC PROSITE; PS00262; G-PROTEIN_RECEP_F2_1; 1.
 CC Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 CC Phosphorylation; Multigene family; G-protein coupled receptor.
 CC NON TER 1 1
 CC TRANSMEM <1 19 1 (POTENTIAL).
 CC DOMAIN 20 33 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 34 54 2 (POTENTIAL).
 FT DOMAIN 55 71 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 72 93 3 (POTENTIAL).
 FT DOMAIN 94 113 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 114 136 4 (POTENTIAL).
 FT DOMAIN 137 158 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 159 181 5 (POTENTIAL).
 FT DOMAIN 182 362 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 363 383 6 (POTENTIAL).
 FT DOMAIN 384 397 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 398 417 7 (POTENTIAL).
 FT DOMAIN 418 440 CYTOPLASMIC (POTENTIAL).
 FT DISULFID 70 150 BY SIMILARITY.
 FT MOD RES 420 420 PHOSPHORYLATION (POTENTIAL).
 FT MOD RES 424 424 PHOSPHORYLATION (POTENTIAL).
 FT MOD RES 439 439 PHOSPHORYLATION (POTENTIAL).
 SQ SEQUENCE 440 AA; 48854 MW; 12B0324E13D37DDF CRC64;

Query Match 47.5%; Score 47; DB 1; Length 440;
 Best Local Similarity 58.8%; Pred. No. 5.8;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
 :||:|||||
 Db 118 IAAWVLSFILMAPAIL 134

RESULT 6
 AC2M2_CHICK STANDARD; PRT; 466 AA.
 AC P30372;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Muscarinic acetylcholine receptor M2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=91373358; PubMed=1840593;
 RA Tietje K.M., Nathanson N.M.;
 RT "Embryonic chick heart expresses multiple muscarinic acetylcholine
 RT receptor subtypes. Isolation and characterization of a gene encoding
 RT a novel m2 muscarinic acetylcholine receptor with high affinity for
 RT pirenzepine.";
 RL J. Biol. Chem. 266:17382-17387(1991).
 CC -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
 CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
 CC BREAKDOWN OF PHOSPHOLIPIDS & MODULATION OF POTASSIUM CHANNELS
 CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 CC ADENYLATE CYCLASE INHIBITION.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- MISCELLANEOUS: This receptor has a high affinity for pirenzepine.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M73217; AAB04106.1;
 CC InterPro; IPR000276; GPCR_Rhodpsn.
 CC Pfam; PF00001; 7tm.1; 1.
 CC PRINTS; PR00237; GPCR_Rhodopsin.
 CC PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.
 CC PROSITE; PS00262; G-PROTEIN_RECEP_F2_1; 1.
 CC PROSITE; PS00262; G-PROTEIN_RECEP_F1_2; 1.

KW Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 KW Phosphorylation; Multigene family; G-protein coupled receptor.
 FT DOMAIN 1 25 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 26 48 1 (POTENTIAL).
 FT DOMAIN 49 62 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 63 83 2 (POTENTIAL).
 FT DOMAIN 84 100 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 101 122 3 (POTENTIAL).
 FT DOMAIN 123 142 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 143 165 4 (POTENTIAL).
 FT DOMAIN 166 187 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 188 210 5 (POTENTIAL).
 FT DOMAIN 211 388 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 389 409 6 (POTENTIAL).
 FT DOMAIN 410 423 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 424 443 7 (POTENTIAL).
 FT DOMAIN 444 466 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 99 179 BY SIMILARITY.
 FT MOD_RES 446 446 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 450 450 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 465 465 PHOSPHORYLATION (POTENTIAL).
 SQ SEQUENCE 466 AA; 51565 MW; 2E3D8241D6168995 CRC64;
 Query Match 47.5%; Score 47; DB 1; Length 466;
 Best Local Similarity 58.8%; Pred. No. 6.1;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAHSHVLSFLWTPTVAL 17
 147 IAAAWLSFILWAPAIL 163
 DB
 RESULT 7
 ACM2 HUMAN STANDARD; PRT; 466 AA.
 AC P08172; QPPIX9;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Muscarinic acetylcholine receptor M2.
 GN CHRM2
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97263421; PubMed=3037705;
 RA Bonner T.I., Buckley N.J., Young A.C., Brann M.R.;
 RT "Identification of a family of muscarinic acetylcholine receptor genes.";
 RL Science 237:527-532 (1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=48166632; PubMed=3443095;
 RA Peralta E.G., Ashkenazi A., Winslow J.W., Smith D.H.,
 RA Ramachandran J., Capon D.J.;
 RT "Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors.";
 RL EMOB J. 6:3923-3929 (1987).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 RA Publ H.L. III, Ikeda S.R., Aronstam R.S.;
 RT "cDNA clones of human proteins involved in signal transduction sequenced by the Guthrie cDNA resource center (www.cdna.org).";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 27-466 FROM N.A.

Kitano T., Kobayakawa H., Saitou N.;
 "Silver project";
 submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE, BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS ADENYLATE CYCLASE INHIBITION.
 -!- SUBCELLULAR LOCATION: Integral membrane protein.
 -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 EMBL; M16404; AA51570.1; --
 EMBL; X15264; CA33335.1; --
 EMBL; AF498916; AA18939.1; --
 EMBL; AB041391; BAA94476.1; --
 PIR; S10126; S10126.
 PDB; 1LUB; 19-JUN-02.
 Genew; HGNC:1951; CHRM2.
 MIM; 118493; --
 GO; GO:0005887; C: integral to plasma membrane; TAS.
 GO; GO:0004981; F: muscarinic acetylcholine receptor activity; TAS.
 GO; GO:0007188; P: G-protein signaling, coupled to cAMP nucle. . .; TAS.
 GO; GO:0007207; P: muscarinic acetyl choline receptor, phospho. . .; TAS.
 GO; GO:0007399; P: neurogenesis; TAS.
 GO; GO:0009405; P: pathogenesis; TAS.
 GO; GO:0008016; P: regulation of heart rate; TAS.
 InterPro; IPR00276; GPCR_Rhodopsn.
 Pfam; PF00001; 7tm.1; 1.
 PRINTS; PR00237; GPCRHOPOSN
 PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
 PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
 KW Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 KW Phosphorylation; Multigene family; G-protein coupled receptor;
 3D-structure.
 FT DOMAIN 1 22 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 23 45 1 (POTENTIAL).
 FT DOMAIN 46 59 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 60 80 2 (POTENTIAL).
 FT DOMAIN 81 97 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 98 119 3 (POTENTIAL).
 FT DOMAIN 120 139 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 140 162 4 (POTENTIAL).
 FT DOMAIN 163 184 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 185 207 5 (POTENTIAL).
 FT DOMAIN 208 388 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 389 409 6 (POTENTIAL).
 FT DOMAIN 410 423 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 424 443 7 (POTENTIAL).
 FT DOMAIN 444 466 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 2 2 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3 3 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 6 6 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 9 9 BY SIMILARITY.
 FT DISULFID 96 176 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 446 446 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 450 450 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 465 465 PHOSPHORYLATION (POTENTIAL).
 SQ SEQUENCE 466 AA; 51715 MW; 2FC2FD7748C2B8C CRC64;
 Query Match 47.5%; Score 47; DB 1; Length 466;
 Best Local Similarity 58.8%; Pred. No. 6.1;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAHSHVLSFLWTPTVAL 17
 :|||:|||||


```

FT TRANSMEM 185 207 5 (POTENTIAL).
FT DOMAIN 208 388 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 389 409 6 (POTENTIAL).
FT DOMAIN 410 423 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 424 443 7 (POTENTIAL).
FT DOMAIN 444 466 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 2 2 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3 3 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 6 6 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DISULFID 96 176 BY SIMILARITY.
FT MOD RES 446 446 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 450 450 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 465 465 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 330 330 K -> N (IN REF. 2).
SQ SEQUENCE 466 AA; 51673 MW; 53D089F179789CD9 CRC64;

Query Match 47.5%; Score 47; DB 1; Length 466;
Best Local Similarity 58.8%; Pred. No. 6.1;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAHSVLSFLWTPYAL 17
Db 144 IAAWVLSFILWAPAIL 160

RESULT 10
ACM2 RAT
ID ACM2 RAT STANDARD; PRT; 466 AA.
AC P10980; Q92221;
DT 01-JUL-1989 (Rel. 11, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Muscarinic acetylcholine receptor M2.
GN CHRM2 OR CHRM-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8068581; PubMed=2825184;
RA Gocayne J.D., Robinson D.A., Fitzgerald M.G., Chung F.-Z.,
RA Kerlavage A.R., Lentes K.-U., Lai J., Wang C.-D., Fraser C.M.,
RA Venter J.C.;
RT "Primary structure of rat cardiac beta-adrenergic and muscarinic
RT cholinergic receptors obtained by automated DNA sequence analysis:
RT further evidence for a multigene family.";
RN Proc. Natl. Acad. Sci. U.S.A. 84:8296-8300(1987).
[2]

SEQUENCE FROM N.A.
TISSUE=Iris;
MEDLINE=99138467; PubMed=9972520;
RA Furuta M., Ohya S., Imaizumi Y., Watanabe M.;
RT "Molecular cloning of m3 muscarinic acetylcholine receptor in rat
RT iris.";
RN J. Smooth Muscle Res. 34:111-122(1998).
CC -1- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE,
CC BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS
CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
CC ADENYLATE CYCLASE INHIBITION.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J03025; AAA40926.1; -.

```

```

DR EMBL; AB017655; BAA36838.1; -.
DR PIR; S10856; S10856.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCRHHODOPSN.
DR PROSITE; PS00237; G_PROTEIN_RECP_F1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECP_F2_1; 1.
KW Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
KW Phosphorylation; Multigene family; G-protein coupled receptor.
FT DOMAIN 1 22 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 23 45 1 (POTENTIAL).
FT DOMAIN 46 59 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 60 80 2 (POTENTIAL).
FT DOMAIN 81 97 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 98 119 3 (POTENTIAL).
FT DOMAIN 120 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 162 4 (POTENTIAL).
FT DOMAIN 163 184 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 185 207 5 (POTENTIAL).
FT DOMAIN 208 388 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 389 409 6 (POTENTIAL).
FT DOMAIN 410 423 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 424 443 7 (POTENTIAL).
FT DOMAIN 444 466 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 2 2 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3 3 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 6 6 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DISULFID 96 176 BY SIMILARITY.
FT MOD RES 446 446 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 450 450 PHOSPHORYLATION (POTENTIAL).
FT MOD RES 465 465 PHOSPHORYLATION (POTENTIAL).
FT CONFLICT 51 51 N -> S (IN REF. 1).
FT CONFLICT 273 273 N -> T (IN REF. 1).
FT CONFLICT 289 290 VS -> SA (IN REF. 1).
FT CONFLICT 313 313 G -> D (IN REF. 1).
FT CONFLICT 337 337 C -> Y (IN REF. 1).
FT CONFLICT 353 353 N -> S (IN REF. 1).
FT CONFLICT 360 360 I -> V (IN REF. 1).
FT CONFLICT 369 369 T -> P (IN REF. 1).
SQ SEQUENCE 466 AA; 51539 MW; 70ECCD86366A676B CRC64;

Query Match 47.5%; Score 47; DB 1; Length 466;
Best Local Similarity 58.8%; Pred. No. 6.1;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 MAHSVLSFLWTPYAL 17
Db 144 IAAWVLSFILWAPAIL 160

RESULT 11
ACM4 RAT
ID ACM4 RAT STANDARD; PRT; 478 AA.
AC P08485;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Muscarinic acetylcholine receptor M4.
GN CHRM4 OR CHRM-4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87263421; PubMed=3037705;
RA Bonner T.I., Buckley N.J., Young A.C., Brann M.R.;
RT "Identification of a family of muscarinic acetylcholine receptor
RT genes.";
RN Science 237:527-532(1987).
CC -1- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
CC CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE.
CC BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS

```

THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 INHIBITION OF ADENYLATE CYCLASE
 -!- SUBCELLULAR LOCATION: Integral membrane protein.
 -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: M16405; AAA51571.1; ALT SEQ.
 InterPro: IPR000276; GPCR_Rhodpsn.
 Pfam: PF00001; 7tm_1; 1.
 PRINTS; PR00237; GPCRHOPOPSN.
 PROSITE; PS00237; G PROTEIN RECEPTOR FL1; 1.
 PROSITE; PS0262; G PROTEIN RECEPTOR FL2; 1.
 Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 Phosphorylation; Multigene family; G-protein coupled receptor.
 DOMAIN 1 30 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 31 53 1 (POTENTIAL).
 DOMAIN 54 67 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 68 88 2 (POTENTIAL).
 DOMAIN 89 105 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 106 127 3 (POTENTIAL).
 DOMAIN 128 147 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 148 170 4 (POTENTIAL).
 DOMAIN 171 192 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 193 215 5 (POTENTIAL).
 DOMAIN 216 400 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 401 421 6 (POTENTIAL).
 DOMAIN 422 435 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 436 455 7 (POTENTIAL).
 DOMAIN 456 478 CYTOPLASMIC (POTENTIAL).
 CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 DISULFID 13 13 BY SIMILARITY.
 MOD_RES 104 184 PHOSPHORYLATION (POTENTIAL).
 MOD_RES 458 458 PHOSPHORYLATION (POTENTIAL).
 MOD_RES 462 462 PHOSPHORYLATION (POTENTIAL).
 MOD_RES 476 476 PHOSPHORYLATION (POTENTIAL).
 SEQUENCE 478 AA; 52920 MW; 2551FF5F7B1A56C CRC64;
 Query Match 46.5%; Score 46; DB 1; Length 478;
 Best Local Similarity 58.8%; Pred. No. 9;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAHSVLSFLWTFYAL 17
 :|||:|||||:
 Db 152 IAAAVLSFLWAPAIL 168
 RESULT 12
 ACW4 HUMAN STANDARD; PRT; 479 AA.
 ID ACW4 HUMAN STANDARD; PRT; 479 AA.
 AC P08173;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Muscarinic acetylcholine receptor M4.
 GN CHRM4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87263421; PubMed=3037705;
 RA Bonner T.I., Buckley N.J., Young A.C., Brann M.R.;
 RA "Identification of a family of muscarinic acetylcholine receptor
 genes.";
 RT Science 237:527-532(1987).

RN REVISIONS TO 258 AND 261-266.
 RP MEDLINE=90166521; PubMed=3272174;
 RX Bonner T.I., Young A.C., Brann M.R., Buckley N.J.;
 RA "Cloning and expression of the human and rat m5 muscarinic
 acetylcholine receptor genes.";
 RL Neuron 1:403-410(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88166632; PubMed=3443095;
 RA Peralta E.G., Ashkenazi A., Winslow J.W., Smith D.H.,
 RA Ramachandran J., Capon D.J.;
 RA "Distinct primary structures, ligand-binding properties and tissue-
 specific expression of four human muscarinic acetylcholine
 receptors.";
 RL EMBO J. 6:3923-3929(1987).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 RC Puhl H.L. III, Ikeda S.R., Aronstam R.S.;
 RA "cDNA clones of human proteins involved in signal transduction
 sequenced by the Guthrie cDNA resource center (www.cdna.org).";
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: The muscarinic acetylcholine receptor mediates various
 cellular responses, including inhibition of adenylate cyclase,
 breakdown of phosphoinositides and modulation of potassium
 channels through the action of G proteins. Primary transducing
 effect is inhibition of adenylate cyclase.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: M16405; AAA51571.1; ALT_SEQ.
 EMBL: X15265; CA33336.1; --
 EMBL: AF498918; AAM18941.1; --
 PIR: S10127; S10127.
 DR HSP; P02699; 1BOJ.
 DR MIM; 118495; --
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0004981; F:muscarinic acetylcholine receptor activity; TAS.
 DR GO; GO:0007213; P:acetyl choline receptor signaling, muscarin. . .; TAS.
 DR GO; GO:0008283; P:cell proliferation; TAS.
 DR GO; GO:0007166; P:cell surface receptor linked signal transdu. . .; TAS.
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR Pfam: PF00001; 7tm_1; 1.
 DR PRINTS; PR00237; GPCRHOPOPSN.
 DR PROSITE; PS00237; G PROTEIN RECEPTOR FL1; 1.
 DR PROSITE; PS0262; G PROTEIN RECEPTOR FL2; 1.
 KW Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;
 KW Phosphorylation; Multigene family; G-protein coupled receptor.
 DOMAIN 1 31 1 (POTENTIAL).
 TRANSMEM 32 54 1 (POTENTIAL).
 DOMAIN 55 68 2 (POTENTIAL).
 TRANSMEM 69 89 2 (POTENTIAL).
 DOMAIN 90 106 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 107 128 3 (POTENTIAL).
 DOMAIN 129 148 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 149 171 4 (POTENTIAL).
 DOMAIN 172 193 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 194 216 5 (POTENTIAL).
 DOMAIN 217 401 CYTOPLASMIC (POTENTIAL).
 TRANSMEM 402 422 6 (POTENTIAL).
 DOMAIN 423 436 EXTRACELLULAR (POTENTIAL).
 TRANSMEM 437 456 7 (POTENTIAL).
 DOMAIN 457 479 CYTOPLASMIC (POTENTIAL).

FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 13 13 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 105 185 BY SIMILARITY.
 FT MOD RES 459 459 PHOSPHORYLATION (POTENTIAL).
 FT MOD RES 463 463 PHOSPHORYLATION (POTENTIAL).
 FT MOD RES 477 477 PHOSPHORYLATION (POTENTIAL).
 FT CONFLICT 184 184 Q -> H (IN REF. 1).
 SQ SEQUENCE 479 AA; 53049 MW; 9CIB05B359A64289 CRC64;
 Query Match 46.5%; Score 46; DB 1; Length 479;
 Best Local Similarity 58.8%; Pred. No. 9;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAHSVLSFLWTPYAL 17
 :|||:|||||:
 Db 153 IAAAVLSFVLWAPAIL 169
 RESULT 13
 ID ACMA_MOUSE STANDARD; PRT; 479 AA.
 AC P32211; Q64036;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Muscarinic acetylcholine receptor M4 (Mm4 mAChr).
 GN CHRM4 OR CHRM-4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93305731; PubMed=7916637;
 RA van Koppen C.J., Lenz W., Nathanson N.M.;
 RT "Isolation, sequence and functional expression of the mouse m4
 muscarinic acetylcholine receptor gene";
 RL Biochim. Biophys. Acta 1173:342-344 (1993).
 [2]
 RP SEQUENCE OF 235-355 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=95179320; PubMed=7874308;
 RA Andre C., Dos Santos G., Koulakoff A.;
 RT "Cultured neurons from mouse brain reproduce the muscarinic receptor
 profile of their tissue of origin.";
 RL Eur. J. Neurosci. 6:1691-1701 (1994).
 CC -!- FUNCTION: THE MUSCARINIC ACETYLCHOLINE RECEPTOR MEDIATES VARIOUS
 CELLULAR RESPONSES, INCLUDING INHIBITION OF ADENYLATE CYCLASE.
 CC BREAKDOWN OF PHOSPHOINOSITIDES & MODULATION OF POTASSIUM CHANNELS
 CC THROUGH THE ACTION OF G PROTEINS. PRIMARY TRANSDUCING EFFECT IS
 CC INHIBITION OF ADENYLATE CYCLASE.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X63473; CAA45071.1; -.
 CC EMBL; S74916; AAB33577.1; -.
 CC PIR; S33776; S33776.
 CC HSSP; P02699; 1BOJ.
 CC MGD; MGI:88399; Chrm4.
 CC InterPro; IPR000276; GPCR_Rhodopsn.
 CC Pfam; PF00001; 7cm_1; 1.
 CC PRINTS; PR00237; GPCRHHODOPSN.
 CC PROSITE; PS00237; G PROTEIN RECEPTOR FL 1; 1.
 CC PROSITE; PS02621; G PROTEIN RECEPTOR FL 2; 1.
 KW Postsynaptic membrane; Ionic channel; Glycoprotein; Transmembrane;

KW Phosphorylation; Multigene family; G-protein coupled receptor.
 FT DOMAIN 1 30 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 31 53 1 (POTENTIAL).
 FT DOMAIN 54 67 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 68 88 2 (POTENTIAL).
 FT DOMAIN 89 105 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 106 127 3 (POTENTIAL).
 FT DOMAIN 128 147 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 148 170 4 (POTENTIAL).
 FT DOMAIN 171 192 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 193 215 5 (POTENTIAL).
 FT DOMAIN 216 401 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 402 422 6 (POTENTIAL).
 FT DOMAIN 423 436 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 437 456 7 (POTENTIAL).
 FT DOMAIN 457 479 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 8 8 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT DISULFID 13 13 BY SIMILARITY.
 FT MOD RES 104 184 PHOSPHORYLATION (POTENTIAL).
 FT MOD RES 459 459 PHOSPHORYLATION (POTENTIAL).
 FT MOD RES 463 463 PHOSPHORYLATION (POTENTIAL).
 FT CONFLICT 322 323 AL -> V (IN REF. 2).
 FT CONFLICT 351 351 S -> N (IN REF. 2).
 SQ SEQUENCE 479 AA; 52973 MW; 04F301E78814CD70 CRC64;
 Query Match 46.5%; Score 46; DB 1; Length 479;
 Best Local Similarity 58.8%; Pred. No. 9;
 Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 MAHSVLSFLWTPYAL 17
 :|||:|||||:
 Db 152 IAAAVLSFVLWAPAIL 168
 RESULT 14
 ID ACMA_CHICK STANDARD; PRT; 490 AA.
 AC P17200;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Muscarinic acetylcholine receptor M4.
 GN CHRM4.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90153912; PubMed=2154460;
 RA Tietje K.M., Goldman P.S., Nathanson N.M.;
 RT "Cloning and functional analysis of a gene encoding a novel
 muscarinic acetylcholine receptor expressed in chick heart and
 brain.";
 RL J. Biol. Chem. 265:2828-2834 (1990).
 CC -!- FUNCTION: The muscarinic acetylcholine receptor mediates various
 CC cellular responses, including inhibition of adenylate cyclase,
 CC breakdown of phosphoinositides and modulation of potassium
 CC channels through the action of G proteins. Primary transducing
 CC effect is inhibition of adenylate cyclase. May couple to multiple
 CC functional responses in cell lines.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Expressed in heart and brain.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:02:48 ; Search time 36 Seconds
(without alignments)
175.288 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 99

Sequence: 1 MAHSLVFLWTPYALKSX 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	51	51.5	355	11	Q9JLY8	Q9JLY8 rattus norv
2	51	51.5	355	11	Q8BVM4	Q8BVM4 mus musculu
3	51	51.5	355	11	Q8RMH9	Q8RMH9 mus musculu
4	50.5	51.0	390	5	Q868G4	Q868G4 brachioisto
5	49	49.5	2124	10	O23057	O23057 arabidopsis
6	47	47.5	440	6	Q9N2A5	Q9N2A5 pongo pygma
7	47	47.5	440	6	Q9N2A7	Q9N2A7 pan troglod
8	47	47.5	456	4	Q9N2A6	Q9N2A6 gorilla gor
9	47	47.5	456	4	Q96RH0	Q96RH0 homo sapien
10	47	47.5	466	11	Q8VH27	Q8VH27 cavia porce
11	47	47.5	495	13	Q801M4	Q801M4 brachydanio
12	46	46.5	110	6	Q98GH6	Q98GH6 oryctolagus
13	46	46.5	166	11	Q88G16	Q88G16 meriones un
14	46	46.5	474	4	Q96RG8	Q96RG8 homo sapien
15	46	46.5	477	11	Q8VH25	Q8VH25 cavia porce
16	46	46.5	487	16	Q836U4	Q836U4 enterococcu

17	46	46.5	1907	16	Q93HI9	Q93HI9 streptomyce
18	45	45.5	333	13	Q91924	Q91924 cynops pyrr
19	45	45.5	355	6	Q9MYJ8	Q9MYJ8 callithrix
20	45	45.5	374	4	Q86VA9	Q86VA9 homo sapien
21	44.5	44.9	51	13	Q9PSF4	Q9PSF4 gallus gall
22	44.5	44.9	310	13	Q91973	Q91973 serinus can
23	44.5	44.9	317	13	Q57447	Q57447 anas platyr
24	44.5	44.9	355	13	Q57449	Q57449 melopaittac
25	44.5	44.9	355	13	Q9W6S1	Q9W6S1 columba liv
26	44.5	44.9	355	13	Q9IA35	Q9IA35 poephila gu
27	44.5	44.9	382	5	Q8T359	Q8T359 platynereis
28	44	44.4	24	8	Q9ZV84	Q9ZV84 melissodes
29	44	44.4	144	17	Q975K5	Q975K5 sulfolobus
30	44	44.4	164	16	Q8EXA9	Q8EXA9 leptospira
31	44	44.4	217	16	Q8G837	Q8G837 bifidobacte
32	44	44.4	301	5	O61895	O61895 caenorhabdi
33	44	44.4	308	16	Q9S2A1	Q9S2A1 streptomyce
34	44	44.4	443	4	Q9H1P0	Q9H1P0 homo sapien
35	44	44.4	544	17	Q97AA6	Q97AA6 thermoplas
36	44	44.4	563	8	Q9ZZA1	Q9ZZA1 trichophyto
37	43.5	43.9	199	16	Q98BY6	Q98BY6 rhizobium 1
38	43.5	43.9	349	13	Q8AYM6	Q8AYM6 brachydanio
39	43.5	43.9	375	13	Q8JID0	Q8JID0 clarias gar
40	43.5	43.9	426	13	Q9PT72	Q9PT72 seriola dum
41	43	43.4	101	4	Q9BZ77	Q9BZ77 homo sapien
42	43	43.4	234	10	Q8LLC2	Q8LLC2 hordeum vul
43	43	43.4	246	16	Q984V1	Q984V1 rhizobium 1
44	43	43.4	313	16	Q7UKS3	Q7UKS3 rhodopirell
45	43	43.4	320	16	Q7VRB2	Q7VRB2 candidatus

ALIGNMENTS

RESULT 1

Q9JLY8 PRELIMINARY; PRT; 355 AA.
AC Q9JLY8
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-JUN-2003 (TREMREL. 24, Last annotation update)
DE Macrophage inflammatory protein-1 alpha receptor.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD.
RX MEDLINE=20555330; PubMed=11091494;
RT Waller A., Naye P., Czaplowski L.G.;
RT "Identification and characterization of a rat macrophage inflammatory
protein-1 alpha receptor."
RL J. Hematother. Stem Cell Res. 9:703-710(2000).
DR EMBL: AF119381; AAF34340.1; -.
DR GO: GO:0016021; C: integral to membrane; IEA.
DR GO: GO:0004872; F: receptor activity; IEA.
DR GO: GO:0001584; F: rhodopsin-like receptor activity; IEA.
DR GO: GO:0007186; P: G-protein coupled receptor protein signalin. .; IEA.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCR_Rhodopsn.
DR PROSITE: PS00237; G PROTEIN RECEPTOR FL1; 1.
DR PROSITE: PS0262; G PROTEIN RECEPTOR FL2; 1.
SQ SEQUENCE 355 AA; 40838 MW; 2FEB8661D1B6E075 CRC64;
Query Match 51.5%; Score 51; DB 11; Length 355;
Best Local Similarity 66.7%; Pred. No. 1.9;
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 3 AHSVLSFLWTPYAL 17
| : : | | | | | | | |


```
DE BAC IG005110.
GN A.IG005110.24 OR F5110.24 OR AT4G00450.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Andrews S.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Waterston R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Wilson R., Andrews S.;
RL "The sequence of A. thaliana F5110.";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA WashU;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Waterston R.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF013293; AAB62842.1; -
DR EMBL; AF195115; AAF02800.1; -
DR EMBL; AL161472; CAB80854.1; -
DR PIR; T01526; T01526.
KW Hypothetical protein.
SQ SEQUENCE 2124 AA; 235154 MW; 84892F6BEF3B65D5 CRC64;
Query Match 49.5%; Score 49; DB 10; Length 2124;
Best Local Similarity 41.2%; Pred. No. 26;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
QY 3 AHSVLSFLWTPTALKS 19
Db 1438 AYSIIAFVLRPFVNS 1454
RESULT 6
Q9N2A5 PRELIMINARY; PRT; 440 AA.
AC Q9N2A5;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Muscarinic acetylcholine receptor m2 (Fragment).
GN CHRM2.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=oran-ul;
RA Kitano T., Kobayakawa H., Saitou N.;
RT "Silver Project.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AB041394; BAA94479.1; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G-PROTEIN RECP_F1_1; 1.
DR PROSITE; PS0262; G-PROTEIN RECP_F1_2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 440 AA; 48870 MW; BDB2FEBE1AC3B83E CRC64;
Query Match 47.5%; Score 47; DB 6; Length 440;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAHSVLSFLWTPTAL 17
Db 118 IAAAWLSFILWAPAIL 134
RESULT 7
Q9N2A7 PRELIMINARY; PRT; 440 AA.
AC Q9N2A7;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Muscarinic acetylcholine receptor m2 (Fragment).
GN CHRM2.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=chimp-220;
RA Kitano T., Kobayakawa H., Saitou N.;
RT "Silver Project.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AB041392; BAA94477.1; -
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G-PROTEIN RECP_F1_1; 1.
DR PROSITE; PS0262; G-PROTEIN RECP_F1_2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON_TER 1
SQ SEQUENCE 440 AA; 48854 MW; 12B0324E13D37DDF CRC64;
Query Match 47.5%; Score 47; DB 6; Length 440;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 1 MAHSVLSFLWTPTAL 17
Db 118 IAAAWLSFILWAPAIL 134
```

```

RESULT 8
Q9N2A6 PRELIMINARY; PRT; 440 AA.
ID Q9N2A6
AC Q9N2A6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscarinic acetylcholine receptor m2 (Fragment).
GN CHRM2.
OS Gorilla gorilla (gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
OX NCBI_TaxID=9593;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=gorilla-UL;
RA Kitano T., Kobayakawa H., Saitou N.;
RT "Silver Project.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DE EMBL; AB041393; BA94478.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL 1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL 2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON TER 1
SQ SEQUENCE 440 AA; 48954 MW; 12B0324E13D37DDF CRC64;

Query Match 47.5%; Score 47; DB 6; Length 440;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
Db 118 IAAAVLSFILWAPAIL 134
: || |||||: ||

RESULT 9
Q96RH0 PRELIMINARY; PRT; 456 AA.
ID Q96RH0
AC Q96RH0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE M2 muscarinic cholinergic receptor.
GN CHRM2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA Aredondo J., Grando S.A.;
RT "Cloning Cholinergic Receptors in Human Keratinocytes.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DE EMBL; AF385588; AA68113.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.

```

```

DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL 1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL 2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
SQ SEQUENCE 456 AA; 50570 MW; F0D02A518097383E CRC64;

Query Match 47.5%; Score 47; DB 4; Length 456;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
Db 144 IAAAVLSFILWAPAIL 160
: || |||||: ||

RESULT 10
Q8VH27 PRELIMINARY; PRT; 466 AA.
ID Q8VH27
AC Q8VH27;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscarinic receptor 2 (Fragment).
GN GPM2.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
[1]
RP SEQUENCE FROM N.A.
RA So I., Yang D., Kim H., Min K., Kim S., Kim K., Park K., Choi K.,
RA Kim I.;
RT "Five subtypes of muscarinic receptors are expressed in gastric smooth
muscles of guinea pig.";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DE EMBL; AY072059; AAL67910.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. . .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G-PROTEIN RECEPTOR FL 1; 1.
DR PROSITE; PS0262; G-PROTEIN RECEPTOR FL 2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON TER 466
SQ SEQUENCE 466 AA; 51630 MW; A992CB6766A37BD5 CRC64;

Query Match 47.5%; Score 47; DB 11; Length 466;
Best Local Similarity 58.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYAL 17
Db 144 IAAAVLSFILWAPAIL 160
: || |||||: ||

RESULT 11
Q801M4 PRELIMINARY; PRT; 495 AA.
ID Q801M4
AC Q801M4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE M2 muscarinic receptor.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
[1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=2229678; PubMed=12411408;
RA Hsieh D.J., Liao C.F.;
RT "Zebrafish M(2) muscarinic acetylcholine receptor: cloning,
RT pharmacological characterization, expression patterns and roles in
RT embryonic bradycardia.";
RL Br. J. Pharmacol. 137:782-792(2002).
DR EMBL: AY039653; AAK93793.1; -.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO: GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCRHHODOPSN.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS0262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 495 AA; 54501 MW; AE98F599B5C3348F CRC64;

Query Match 47.5%; Score 47; DB 13; Length 495;
Best Local Similarity 58.8%; Pred. No. 13;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHVSLSFLLWTPYAL 17
   :|||:|||||
Db 152 IAAAWLSFLLWAPAIL 168

RESULT 12
ID Q9BGH6 PRELIMINARY; PRT; 110 AA.
AC Q9BGH6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Putative muscarinic receptor M4 (Fragment).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Sclera, and Liver;
RA Thu M.K., Beuterman R.W., Liang J.Z.;
RT "Differential expression of muscarinic receptors in postnatal rabbit
RT sclera.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL: AF308570; AAG49559.1; -.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO: GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCRHHODOPSN.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS0262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON_TER 1
FT NON_TER 110
SQ SEQUENCE 110 AA; 12444 MW; C47A29DDB13CA00B CRC64;

Query Match 46.5%; Score 46; DB 6; Length 110;
Best Local Similarity 58.8%; Pred. No. 4.3;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHVSLSFLLWTPYAL 17
   :|||:|||||
Db 43 IAAAWLSFLLWAPAIL 59

```

```

RESULT 13
ID Q88616 PRELIMINARY; PRT; 166 AA.
AC Q88616;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Muscarinic acetylcholine receptor M4 (Fragment).
GN CHRM4.
OS Meriones unguiculatus (Mongolian jird) (Mongolian gerbil).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Gerbillinae;
OC Meriones.
OX NCBI_TaxID=10047;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21431793; PubMed=11547340;
RA Wangemann P., Liu J., Scherer E.Q., Herzog M., Shimozono M.,
RA Scofield M.A.;
RT "Muscarinic receptors control K+ secretion in inner ear strial
RT marginal cells.";
RL J. Membr. Biol. 182:171-181(2001).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL: AF079114; AAC28098.1; -.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO: GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro: IPR000276; GPCR_Rhodopsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCRHHODOPSN.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS0262; G_PROTEIN_RECEP_F1_2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
FT NON_TER 1
FT NON_TER 166
SQ SEQUENCE 166 AA; 18477 MW; B13F1DC7B73DF41F CRC64;

Query Match 46.5%; Score 46; DB 11; Length 166;
Best Local Similarity 58.8%; Pred. No. 6.5;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 MAHVSLSFLLWTPYAL 17
   :|||:|||||
Db 120 IAAAWLSFLLWAPAIL 136

RESULT 14
ID Q96RG8 PRELIMINARY; PRT; 474 AA.
AC Q96RG8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE M4 muscarinic cholinergic receptor.
GN CHRM4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Arredondo J., Grando S.A.;
RT "Cloning Cholinergic Receptors in Human Keratinocytes.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL: AF385590; AAK68115.1; -.
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:07:01 ; Search time 54 Seconds
(without alignments)
104.647 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHVSLSFLWTPYALKSX 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19	95.0	19	6	ADA40708	ADA40708 Human sec
2	19	95.0	19	6	ABR47713	ABR47713 Human sec
3	19	95.0	20	2	RAY07853	RAY07853 Human sec
4	7	35.0	57	5	ABP32031	ABP32031 Human ORF
5	7	35.0	113	4	AAW83410	AAW83410 Human imm
6	7	35.0	285	4	ABW79589	ABW79589 Coryneb
7	7	35.0	355	2	AAW29179	AAW29179 Rat CC ch
8	7	35.0	355	7	ADD45360	ADD45360 Rat Prote
9	7	35.0	1019	4	AAW2745	AAW2745 C glutami
10	7	35.0	1019	5	ABB79444	ABB79444 Coryneb
11	6	30.0	33	2	AAW60093	AAW60093 Human end
12	6	30.0	33	4	ABW64900	ABW64900 Human sec
13	6	30.0	33	4	ABW64424	ABW64424 Human sec
14	6	30.0	38	3	AAW24890	AAW24890 Arabidops
15	6	30.0	41	3	ABW44612	ABW44612 Human sec
16	6	30.0	41	4	AAW06857	AAW06857 Human foe
17	6	30.0	41	4	AAW06447	AAW06447 Human foe
18	6	30.0	55	5	ABP00846	ABP00846 Human ORF
19	6	30.0	59	2	AAW68891	AAW68891 Japanese
20	6	30.0	59	3	AAW58213	AAW58213 Arabidops
21	6	30.0	62	5	ABP35008	ABP35008 Human ORF
22	6	30.0	63	3	AAW29813	AAW29813 Human sec
23	6	30.0	63	3	ABW99744	ABW99744 Human sec
24	6	30.0	63	6	ABR01235	ABR01235 Human gen
25	6	30.0	63	7	ADC20501	ADC20501 Human sec

ALIGNMENTS

RESULT 1

ADA40708
ID ADA40708 standard; protein; 19 AA.

XX AC ADA40708;

XX XX 20-NOV-2003 (first entry)

XX DE Human secreted protein.

XX XX Human; secreted protein; cancer; hyperproliferative disorder;

KW rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;

KW anaemia; allergic reaction; asthma; cardiovascular disorder;

KW wound healing; cytostatic; immunosuppressive; neoplastic; neuroprotective;

KW antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory;

KW vulnery; cardiant; gene therapy.

XX OS Homo sapiens.

XX XX WO2002102993-A2.

XX XX 27-DEC-2002.

XX XX 19-MAR-2002; 2002WO-US008123.

XX XX 21-MAR-2001; 2001US-0277340P.

XX XX 19-JUL-2001; 2001US-0306171P.

XX XX 13-NOV-2001; 2001US-0331287P.

XX XX (HUMA-) HUMAN GENOME SCI INC.

XX DI Rosen CA, Ruben SM;

XX XX WPI; 2003-175238/17.

XX XX New human secreted proteins and nucleic acid molecules, useful for preparing a diagnostic or pharmaceutical composition for diagnosing, preventing or treating cancer or other hyperproliferative disorder, asthma, allergies or AIDS.

XX PS Claim 1; SEQ ID NO 1090; 3205pp; English.

XX CC The invention relates to novel genes ADA39629-ADA40565 and proteins ADA40566-ADA41501 for human secreted proteins, useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The polypeptides, nucleic acid molecules, antibodies or their fragments, and agonists or antagonists that bind to the polypeptide are useful for preparing a diagnostic or pharmaceutical composition for

Aam99722 Human exc
Aam42337 Human kid
Aga40922 Human sec
Adc74224 Human sec
Aay02749 Human sec
Ada07428 Human sec
Aam81705 Human hae
Aam81448 Human hae
Aam00848 Human bon
Aau51440 Propionib
Aam47959 Propionib
Aao11233 Human pol
Aam94382 Human rep
Aau65663 Propionib
Aam62182 Propionib
Aay87138 Human sec
Aae06115 Human gen
Abg333937 Human sec
Abp07585 Human ORF
Aao11478 Human pol

CC diagnosing or treating cancer or other hyperproliferative disorder. The
 CC polypeptides and nucleic acid molecules are also useful for detecting,
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer
 CC or other hyperproliferative disorders including neoplasms, autoimmune
 CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic
 CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,
 CC thrombocytopenia), allergic reactions including asthma or eczema,
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,
 CC fungal or viral infections including HIV/AIDS), or wound healing and
 CC disorders of epithelial cell proliferation. The nucleic acids are also
 CC useful for chromosome identification, radiation hybrid mapping or long-
 CC range restriction mapping, as molecular weight markers, or as
 CC hybridization or diagnostic probes. The polypeptides and antibodies are
 CC useful for providing immunological probes for differential identification
 CC of the tissues immunohistochemistry assays. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 19 AA;

Query Match 95.0%; Score 19; DB 6; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYALKS 19
 Db 1 MAHSVLSFLWTPYALKS 19

RESULT 2
 ABR47713
 ID ABR47713 standard; protein; 19 AA.

XX AC ABR47713;

XX DT 12-JUN-2003 (first entry)

XX DE Human secreted protein, SEQ ID 604.

XX KW Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;
 KW vulnaray; antiinflammatory; nootropic; neuroprotective;
 KW antiparkinsonian; gene therapy; human; cardiovascular disorder.

XX OS Homo sapiens.

XX PN WO200295010-A2.

XX PD 28-NOV-2002.

XX PF 19-MAR-2002; 2002WO-US009785.

XX PR 21-MAR-2001; 2001US-0277340P.

XX PR 19-JUL-2001; 2001US-0306171P.

XX PR 13-NOV-2001; 2001US-0331287P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX DR WPI; 2003-129429/12.

XX Novel human secreted proteins, useful for detecting, preventing,
 PT diagnosing, prognosticating, treating and/or ameliorating cardiovascular
 PT disorders such as arrhythmia.

XX Claim 13; SEQ ID NO 604; 188pp; English.

XX The present invention relates to novel human secreted proteins (ABR47633-

CC ABR48145) and their coding sequences (ACC50344-ACC50856). The proteins
 CC and their coding sequences are useful for the preparation of a diagnostic
 CC or pharmaceutical composition for diagnosing or treating a cardiovascular
 CC disorder (e.g., arrhythmia, tachycardia, cardiac arrest, coronary
 CC arteriosclerosis and myocardial ischaemia), neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, renal disorders,
 CC proliferative disorders and/or cancerous diseases and conditions, for
 CC wound healing and epithelial cell proliferation, to treat inflammation or
 CC infection, for treating thrombosis and arteriosclerosis, for treating or
 CC preventing neural damage which occurs in neuronal disorders or
 CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's
 CC disease, to enhance bone and periodontal regeneration and aid in tissue
 CC transplants or bone grafts, to prevent skin aging or hair loss, to
 CC stimulate growth and differentiation of haematopoietic cells and bone
 CC marrow cells when used in combination with other cytokines, to maintain
 CC organs before transplantation or for supporting cell culture of primary
 CC tissues, to increase or decrease differentiation or proliferation of
 CC embryonic stem cells, or to modulate mammalian characteristics or
 CC metabolism. Note: The sequence data for this patent was published in
 CC electronic format and is available from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 19 AA;

Query Match 95.0%; Score 19; DB 6; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1e-12;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAHSVLSFLWTPYALKS 19
 Db 1 MAHSVLSFLWTPYALKS 19

RESULT 3
 AAY07853

ID AAY07853 standard; protein; 20 AA.

XX AC AAY07853;

XX DT 06-JUL-1999 (first entry)

XX DE Human secreted protein fragment encoded from gene 2.

XX KW Human; secreted protein; treatment; prevention; protein therapy; AIDS;
 KW gene therapy; diagnosis; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
 KW immune system disease; autoimmune disease; hepatic disease; lymphoma;
 KW renal disease; inflammation; allergy; Alzheimer's disease; schizophrenia;
 KW cognitive disorder; prostate disease; skeletal; cardiac; muscle disorder;
 KW pulmonary disorder; transplant rejection; osteoclast; osteoporosis;
 KW arthritis; malignancy; digestive; infection.

XX OS Homo sapiens.

XX PN WO9918208-A1.

XX PD 15-APR-1999.

XX PF 01-OCT-1998; 98WO-US020775.

XX PR 02-OCT-1997; 97US-0060833P.

XX PR 02-OCT-1997; 97US-0060836P.

XX PR 02-OCT-1997; 97US-0060837P.

XX PR 02-OCT-1997; 97US-0060838P.

XX PR 02-OCT-1997; 97US-0060839P.

XX PR 02-OCT-1997; 97US-0060843P.

XX PR 02-OCT-1997; 97US-0060862P.

XX PR 02-OCT-1997; 97US-0060866P.

XX PR 02-OCT-1997; 97US-0060874P.

XX PR 02-OCT-1997; 97US-0060880P.

XX PR 02-OCT-1997; 97US-0060884P.

DE Human immune/haematopoietic antigen SEQ ID NO:11003.
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
OS Homo sapiens.
XX WO200157182-A2.
PN 09-AUG-2001.
PD XX
XX 17-JAN-2001; 2001WO-US001354.
PF XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184564P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216800P.
PR 11-JUL-2000; 2000US-0217480P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225466P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232197P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-024617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.

PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.
 DR N-ESDB; AAK56191.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX
 PS Claim 11; SEQ ID NO 11003; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 113 AA;

Query Match 35.0%; Score 7; DB 4; Length 113;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SVLSFLL 11
 Db 34 SVLSFLL 40

RESULT 6
 AAB79589
 ID AAB79589 standard; protein; 285 AA.
 XX
 AC AAB79589;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:694.
 XX
 KW Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200100844-A2.
 XX
 XX 04-JAN-2001.
 PD
 XX 23-JUN-2000; 2000WO-IB000943.
 PF
 XX 25-JUN-1999; 99US-0141031P.
 PR 08-JUL-1999; 99DE-01031412.

PR 08-JUL-1999; 99DE-01031413.
 PR 08-JUL-1999; 99DE-01031419.
 PR 08-JUL-1999; 99DE-01031420.
 PR 08-JUL-1999; 99DE-01031424.
 PR 08-JUL-1999; 99DE-01031428.
 PR 08-JUL-1999; 99DE-01031431.
 PR 08-JUL-1999; 99DE-01031433.
 PR 08-JUL-1999; 99DE-01031434.
 PR 08-JUL-1999; 99DE-01031510.
 PR 08-JUL-1999; 99DE-01031562.
 PR 08-JUL-1999; 99DE-01031634.
 PR 09-JUL-1999; 99DE-01032180.
 PR 09-JUL-1999; 99DE-01032227.
 PR 09-JUL-1999; 99DE-01032230.
 PR 09-JUL-1999; 99US-0143208P.
 PR 14-JUL-1999; 99DE-01032924.
 PR 14-JUL-1999; 99DE-01032973.
 PR 14-JUL-1999; 99DE-01033005.
 PR 27-AUG-1999; 99DE-01040765.
 PR 31-AUG-1999; 99US-0151572P.
 PR 03-SEP-1999; 99DE-01042076.
 PR 03-SEP-1999; 99DE-01042079.
 PR 03-SEP-1999; 99DE-01042086.
 PR 03-SEP-1999; 99DE-01042087.
 PR 03-SEP-1999; 99DE-01042088.
 PR 03-SEP-1999; 99DE-01042095.
 PR 03-SEP-1999; 99DE-01042123.
 PR 03-SEP-1999; 99DE-01042125.
 XX
 XX (BADI) BASF AG.
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
 DR WPI; 2001-061975/07.
 DR N-ESDB; AAF71706.
 XX
 XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 FT carbohydrates or enzymes.
 XX
 PS Claim 20; Page 1129-1130; 1246pp; English.
 XX
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243
 CC to AAB 79633 which are involved in carbon metabolism and energy
 CC production. The C. glutamicum SMP gene can be used in vectors (II) for
 CC expression in host cells and production or modulation of production of
 CC fine chemicals, such as, an organic acid, a proteinogenic or
 CC nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a
 CC nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid,
 CC a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a
 CC polypeptide, or an enzyme. The presence of (I) or SMP proteins (III)
 CC encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to C.
 CC glutamicum, identify and localise C. glutamicum sequences of interest, in
 CC evolutionary studies, in determining SMP protein regions required for
 CC function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH)
 XX
 SQ Sequence 285 AA;

Query Match 35.0%; Score 7; DB 4; Length 285;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SVLSFLL 11
 Db 141 SVLSFLL 147

```

RESULT 7
AAW29179
ID AAW29179 standard; protein; 355 AA.
XX
AC AAW29179;
XX
DT 19-DEC-1997 (first entry)
XX
DE Rat CC chemokine receptor.
XX
KW rat; CC chemokine receptor; screen; binding; ligand.
XX
OS Rattus rattus.
XX
PN JP09227599-A.
XX
PD 02-SEP-1997.
XX
PF 22-FEB-1996; 96JP-00035192.
XX
PR 22-FEB-1996; 96JP-00035192.
XX
PA (TAKE ) TAKEDA CHEM IND LTD.
XX
DR WPI; 1997-486426/45.
XX
DR N-PSDB; AAT86839.
XX
CC chemokine receptor protein - useful to screen for novel binding
PT compounds.
XX
Claim 1; Page 20-21; 26pp; Japanese.
XX
This sequence is a rat CC chemokine receptor. The receptor can be used to
CC screen for novel binding compounds and for preparation of antibodies or
CC antiserum
XX
SQ Sequence 355 AA;
Query Match 35.0%; Score 7; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLIWTPT 15
Db 249 FLIWTPT 255
|||||

RESULT 8
ADD45360
ID ADD45360 standard; protein; 355 AA.
XX
AC ADD45360;
XX
DT 29-JAN-2004 (first entry)
XX
DE Rat Protein NP_065417, SEQ ID NO 10793.
XX
KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
XX chronic constriction injury; CCI; spared nerve injury; SNI; Chung.
XX
OS Rattus norvegicus.
XX
PN WO2003016475-A2.
XX
PD 27-FEB-2003.
XX
PF 14-AUG-2002; 2002WO-US025765.
XX
PR 14-AUG-2001; 2001US-0312147P.
XX
PR 01-NOV-2001; 2001US-0346382P.
XX
PR 26-NOV-2001; 2001US-0333347P.
XX
PA (GEO ) GEN HOSPITAL CORP.

(PARB ) BAYER AG.
Woelf C, D'urso D, Befort K, Costigan M;
WPI; 2003-268312/26.
GENBANK; NP_065417.
New composition comprising two or more isolated polypeptides, useful for
preparing a medicament for treating pain in an animal.
Claim 1; Page; 1017pp; English.
The invention discloses a composition comprising two or more isolated rat
or human polynucleotides or a polynucleotide which represents a fragment,
derivative or allelic variation of the nucleic acid sequence. Also
claimed are a vector comprising the novel polynucleotide, a host cell
comprising the vector, a method for identifying a nucleotide sequence
which is differentially regulated in an animal subjected to pain and a
kit to perform the method, an array, a method for identifying an agent
that increases or decreases the expression of the polynucleotide sequence
that is differentially expressed in neuronal tissue of a first animal
subjected to pain, a method for identifying a compound which regulates
the expression of a polynucleotide sequence which is differentially
expressed in an animal subjected to pain, a method for identifying a
compound that regulates the activity of one or more of the
polynucleotides, a method for producing a pharmaceutical composition, a
method for identifying a compound or small molecule that regulates the
activity in an animal of one or more of the polypeptides given in the
specification, a method for identifying a compound useful in treating
pain and a pharmaceutical composition comprising the one or more
polypeptides or their antibodies. The polynucleotide or the compound that
modulates its activity is useful for preparing a medicament for treating
pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
therapy). The sequence presented is a rat protein (shown in Table 2 of
the specification) which is differentially expressed during pain. Note:
The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic form directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences.
Sequence 355 AA;
Query Match 35.0%; Score 7; DB 7; Length 355;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLIWTPT 15
Db 249 FLIWTPT 255
|||||

RESULT 9
AAG92745
ID AAG92745 standard; protein; 1019 AA.
XX
AC AAG92745;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6499.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-00127688.
XX
PR 16-DEC-1999; 99JP-00377484.

```


XX This invention describes novel human nucleic acid (cDNA) sequences (A),
 CC that are highly expressed in uterine tumour tissue and which have
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
 CC are used (i) to identify agents suitable for treatment of uterine or
 CC endometrial cancer; (ii) directly for treating these forms of cancer
 CC (including expression from gene therapy vectors) and (iii) for generation
 CC of specific antibodies. (A) are identified by assembling ESTs (expressed
 CC sequence tags) from a particular tissue type before comparison of the
 CC expression patterns. This allows a significantly longer fragment of the
 CC gene to be revealed, so should reduce the number of failures associated
 CC with the fact that ESTs from different libraries may represent different
 CC parts of the same unknown gene, distorting the estimated frequency of
 CC occurrence in a particular tissue. AAY59941-Y60328 represent protein
 CC fragments encoded by the human endometrium tumour cDNA library derived
 CC EST fragments represented in AAZ41991-Z42121
 XX
 SQ Sequence 33 AA;

Query Match 30.0%; Score 6; DB 2; Length 33;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
 |||||
 Db 4 VLSFLL 9

RESULT 12

AAB64900
 ID AAB64900 standard; protein; 33 AA.

AC AAB64900;

XX 23-MAR-2001 (first entry)

XX Human secreted protein sequence encoded by gene 19 SEQ ID NO:78.

XX Human; secreted protein; diagnosis; immunomodulatory; anti-sclerotic;
 KW dermatological; immunosuppressive; anti-inflammatory; anti-HIV;
 KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
 KW ophthalmological; neuroprotectant; neurotropic; anticonvulsant; vulnary;
 KW antialzheimers; antiparkinsonian; antimicrobial; immune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; HIV; infection;
 KW hyperproliferative disorder; cancer; Gaucher's disease; wound healing;
 KW cardiovascular disease; Scimitar syndrome; Chaga's cardiomyopathy;
 KW coronary arteriosclerosis; angiogenic disorder; diabetic retinopathy;
 KW corneal graft neovascularisation; neurological disorder; regeneration;
 KW Huntington's chorea; Alzheimer's disease; Parkinson's disease;
 KW infectious disease; chemotaxis.

XX Homo sapiens.

XX WO200076530-A1.

XX 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US014933.

XX 11-JUN-1999; 99US-0138572P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (ROSE/) ROSEN C A.

XX Rosen CA, Ruben SM, Komatsoulis GA;

XX WPI; 2001-071147/08.

XX N-PSDB; AAF333231.

XX Nucleic acids encoding 49 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's disease
 PT and diabetic retinopathy.

XX Claim 11; Page 494; 554pp; English.
 XX The polynucleotide sequences given in AAF33213 to AAF33261 encode the
 CC human secreted proteins given in AAB64882 to AAB64930. AAB64931 to
 CC AAB64991 represent human secreted polypeptide sequences and proteins
 CC homologous to them, which are given in the exemplification of the present
 CC invention. Human secreted proteins have activities based on the tissues
 CC and cells the genes are expressed in. Examples of activities include:
 CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;
 CC anti-inflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
 CC vascular; antimicrobial; anti-angiogenic; ophthalmological;
 CC neuroprotectant; anticonvulsant; neurotropic; antialzheimers;
 CC antiparkinsonian; and vulnary. The polynucleotides and polypeptides can
 CC be used in the prevention, diagnosis and treatment of diseases associated
 CC with inappropriate polypeptide expression. Disorders that may be
 CC prevented, diagnosed and/or treated by the above methods include immune
 CC disorders (e.g. multiple sclerosis, systemic lupus erythematosus and
 CC human immuno-deficiency virus (HIV) infections), hyperproliferative
 CC disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
 CC (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
 CC arteriosclerosis), angiogenic disorders (e.g. corneal graft
 CC neovascularisation and diabetic retinopathy), neurological disorders
 CC (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
 CC infectious diseases and/or for promoting wound healing, regeneration and
 CC /or chemotaxis. AAF33204 to AAF33212 and AAB64881 represent sequences
 CC used in the exemplification of the present invention
 XX

SQ Sequence 33 AA;

Query Match 30.0%; Score 6; DB 4; Length 33;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11

Db 3 VLSFLL 8

RESULT 13

AAB64424
 ID AAB64424 standard; protein; 33 AA.

XX AAB64424;

XX 23-MAR-2001 (first entry)

XX Human secreted protein sequence encoded by gene 3 SEQ ID NO:62.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; cerebral ischaemia; infection;
 KW angiogenesis; nervous system disorder; Alzheimer's disease; skin aging;
 KW ocular disorder; corneal infection; wound healing; food additive;
 KW preservative.

XX Homo sapiens.

XX WO200077255-A1.

XX 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US014926.

XX 11-JUN-1999; 99US-0138628P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis GA;

XX

DR WPI; 2001-025337/03.
 DR N-PSDB; AAF32701.
 XX Isolated nucleic acid molecule encoding a human secreted protein is used
 FT in preventing, treating or ameliorating a medical condition.
 XX
 XX
 PS Claim 11; Page 514; 593pp; English.
 XX
 CC The polynucleotide sequences given in AAF32699 to AAF32747 encode the
 CC human secreted proteins given in AAF64422 to AAF64470. AAF64471 to
 CC AAF64548 represent human secreted polypeptide sequences and proteins
 CC homologous to them, which are given in the exemplification of the present
 CC invention. Human secreted proteins have activities based on the tissues
 CC and cells the genes are expressed in. Examples of activities include:
 CC antiarthritic; immunosuppressive; antineumatic; antiproliferative;
 CC cytoskeletal; cardiac; vasotropic; cerebroprotective; nootropic;
 CC neuroprotective; antibacterial; virucide; fungicide; and
 CC ophthalmological. The polynucleotides and polypeptides can be used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathological condition. Disorders which are diagnosed or treated include
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities. AAF32699 to AAF32698 and
 CC AAF64421 represent sequences used in the exemplification of the present
 CC invention
 XX
 SQ Sequence 33 AA;
 Query Match 30.0%; Score 6; DB 4; Length 33;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 VLSFLL 11
 Db 3 VLSFLL 8
 |||||
 |||||
 RESULT 14
 AAG24890
 ID AAG24890 standard; protein; 38 AA.
 XX
 AC AAG24890;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 28737.
 XX
 XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 XX Arabidopsis thaliana.
 OS
 XX
 DN EPI033405-A2.
 XX
 XX 06-SEP-2000.
 PD
 XX
 XX 25-FEB-2000; 2000EP-00301439.
 PF
 XX
 XX 25-FEB-1999; 99US-0121825P.
 PR
 PR 05-MAR-1999; 99US-0123180P.
 PR
 PR 09-MAR-1999; 99US-0123548P.
 PR
 PR 23-MAR-1999; 99US-0125788P.
 PR
 PR 25-MAR-1999; 99US-0126264P.
 PR
 PR 29-MAR-1999; 99US-0126785P.
 PR
 PR 01-APR-1999; 99US-0127462P.
 PR
 PR 06-APR-1999; 99US-0128234P.
 PR
 PR 08-APR-1999; 99US-0128714P.
 PR
 PR 16-APR-1999; 99US-0129845P.
 PR
 PR 19-APR-1999; 99US-0130077P.
 PR
 PR 21-APR-1999; 99US-0130449P.
 PR
 PR 23-APR-1999; 99US-0130510P.
 PR
 PR 28-APR-1999; 99US-0130891P.
 PR
 PR 30-APR-1999; 99US-0131449P.
 PR
 PR 30-APR-1999; 99US-0132048P.
 PR
 PR 30-APR-1999; 99US-0132407P.
 PR
 PR 04-MAY-1999; 99US-0132484P.
 PR
 PR 05-MAY-1999; 99US-0132485P.
 PR
 PR 06-MAY-1999; 99US-0132486P.
 PR
 PR 07-MAY-1999; 99US-0132487P.
 PR
 PR 11-MAY-1999; 99US-0134256P.
 PR
 PR 14-MAY-1999; 99US-0134218P.
 PR
 PR 14-MAY-1999; 99US-0134219P.
 PR
 PR 14-MAY-1999; 99US-0134221P.
 PR
 PR 18-MAY-1999; 99US-0134370P.
 PR
 PR 19-MAY-1999; 99US-0134768P.
 PR
 PR 20-MAY-1999; 99US-0134941P.
 PR
 PR 20-MAY-1999; 99US-0135124P.
 PR
 PR 21-MAY-1999; 99US-0135353P.
 PR
 PR 24-MAY-1999; 99US-0135629P.
 PR
 PR 25-MAY-1999; 99US-0136021P.
 PR
 PR 27-MAY-1999; 99US-0136392P.
 PR
 PR 28-MAY-1999; 99US-0136782P.
 PR
 PR 01-JUN-1999; 99US-0137222P.
 PR
 PR 03-JUN-1999; 99US-0137528P.
 PR
 PR 04-JUN-1999; 99US-0137502P.
 PR
 PR 07-JUN-1999; 99US-0137724P.
 PR
 PR 08-JUN-1999; 99US-0138094P.
 PR
 PR 10-JUN-1999; 99US-0138540P.
 PR
 PR 10-JUN-1999; 99US-0138847P.
 PR
 PR 14-JUN-1999; 99US-0139119P.
 PR
 PR 16-JUN-1999; 99US-0139452P.
 PR
 PR 17-JUN-1999; 99US-0139453P.
 PR
 PR 17-JUN-1999; 99US-0139492P.
 PR
 PR 18-JUN-1999; 99US-0139454P.
 PR
 PR 18-JUN-1999; 99US-0139455P.
 PR
 PR 18-JUN-1999; 99US-0139456P.
 PR
 PR 18-JUN-1999; 99US-0139457P.
 PR
 PR 18-JUN-1999; 99US-0139458P.
 PR
 PR 18-JUN-1999; 99US-0139459P.
 PR
 PR 18-JUN-1999; 99US-0139460P.
 PR
 PR 18-JUN-1999; 99US-0139461P.
 PR
 PR 18-JUN-1999; 99US-0139462P.
 PR
 PR 18-JUN-1999; 99US-0139463P.
 PR
 PR 18-JUN-1999; 99US-0139750P.
 PR
 PR 18-JUN-1999; 99US-0139763P.
 PR
 PR 21-JUN-1999; 99US-0139817P.
 PR
 PR 22-JUN-1999; 99US-0139899P.
 PR
 PR 23-JUN-1999; 99US-0140353P.
 PR
 PR 23-JUN-1999; 99US-0140354P.
 PR
 PR 24-JUN-1999; 99US-0140695P.
 PR
 PR 28-JUN-1999; 99US-0140823P.
 PR
 PR 29-JUN-1999; 99US-0140991P.
 PR
 PR 30-JUN-1999; 99US-0141287P.
 PR
 PR 01-JUL-1999; 99US-0141842P.
 PR
 PR 01-JUL-1999; 99US-0142154P.
 PR
 PR 02-JUL-1999; 99US-0142055P.
 PR
 PR 06-JUL-1999; 99US-0142390P.
 PR
 PR 08-JUL-1999; 99US-0142803P.
 PR
 PR 09-JUL-1999; 99US-0142920P.
 PR
 PR 12-JUL-1999; 99US-0142977P.
 PR
 PR 13-JUL-1999; 99US-0143542P.
 PR
 PR 14-JUL-1999; 99US-0143624P.
 PR
 PR 15-JUL-1999; 99US-0144005P.
 PR
 PR 16-JUL-1999; 99US-0144085P.

PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145313P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.

PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0158293P.
PR 13-OCT-1999; 99US-0158294P.
PR 13-OCT-1999; 99US-0158295P.
PR 14-OCT-1999; 99US-0158329P.
PR 14-OCT-1999; 99US-0158330P.
PR 14-OCT-1999; 99US-0158331P.
PR 14-OCT-1999; 99US-0158637P.
PR 14-OCT-1999; 99US-0158638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 30.0%; Score 6; DB 3; Length 38;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
|||
Db 32 VLSFLL 37

RESULT 15

AAB44612

ID AAB44612 standard; protein; 41 AA.

XX AAB44612;

XX 12-FEB-2001 (first entry)

XX Human secreted protein sequence encoded by gene 17 SEQ ID NO:77.

Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
antidiabetic; antiinflammatory; antitumor; antiparasitic; anticonvulsant;
antibacterial; antifungal; antiparasitic; cardiac; gene therapy;
food additive; preservative; chromosome identification; cancer;
immune disorder; cardiovascular disorder; neurological disease;
wound healing; infectious disease.

XX Homo sapiens.

XX WO200058339-A2.

XX 05-OCT-2000.

XX 22-MAR-2000; 2000WO-US007440.

XX 26-MAR-1999; 99US-0126503P.

XX 17-DEC-1999; 99US-0172409P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Komatsoulis G;

XX

DR WPI: 2000-594637/56.
DR N-PSDB; AAC79697.
XX
XX Fifty nucleic acid molecules encoding human secreted proteins, useful in
PT the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases.
XX
XX Claim 11; Page 375; 410pp; English.
XX
XX The polynucleotide sequences given in AAC79681 to AAC79730 encode the
CC human secreted proteins given in AAB44596 to AAB44645. AAB44646 to
CC AAB44693 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Examples of activities include:
CC cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
CC antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
CC vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
CC cardiant. The polynucleotides and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides
CC can also be used as a food additive or preservative to increase or
CC decrease storage capabilities. The polynucleotides are useful for
CC chromosome identification. They are also useful as probes for diagnosing
CC a disorder related to the female reproductive system, particularly breast
CC and/or ovary cancer. They are also useful in the gene therapy of breast
CC and ovarian cancer. The nucleic acids, protein, antibodies, agonists and
CC antagonists from the present invention are useful in the diagnosis,
CC treatment and prevention of: cancer; immune disorders; cardiovascular
CC disorders; wound healing; neurological diseases; and infectious diseases.
CC AAC79672 to AAC79680 and AAB44595 represent sequences used in the
XX exemplification of the present invention

SQ Sequence 41 AA;

Query Match 30.0%; Score 6; DB 3; Length 41;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LSFLW 12
|
|
|
|
|
|
|
Db 6 LSFLW 11

Search completed: July 19, 2004, 17:13:09
Job time : 57 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:12:07 ; Search time 18 Seconds
(without alignments)
57.362 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pbp.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB.pbp.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB.pbp.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB.pbp.*
- 5: /cgn2_6/ptodata/2/iaa/PTCUS COMB.pbp.*
- 6: /cgn2_6/ptodata/2/iaa/backfile1.pbp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	35.0	355	4	US-09-886-319A-13
2	6	30.0	33	4	US-09-673-395A-293
3	6	30.0	80	4	US-09-227-357-249
4	6	30.0	105	4	US-09-482-273-177
5	6	30.0	124	4	US-09-328-352-6057
6	6	30.0	148	4	US-09-489-039A-11733
7	6	30.0	154	4	US-09-313-458-34
8	6	30.0	154	4	US-09-746-359A-8
9	6	30.0	176	4	US-09-313-458-19
10	6	30.0	176	4	US-09-746-359A-5
11	6	30.0	215	4	US-09-252-991A-18224
12	6	30.0	338	4	US-09-107-532A-6781
13	6	30.0	345	4	US-09-482-273-113
14	6	30.0	369	4	US-09-482-273-208
15	6	30.0	377	1	US-08-153-848-42
16	6	30.0	377	3	US-09-299-843A-42
17	6	30.0	377	4	US-09-088-337B-42
18	6	30.0	377	5	PCT-US93-11153-42
19	6	30.0	408	4	US-09-107-532A-6563
20	6	30.0	426	4	US-09-134-000C-6607
21	6	30.0	431	4	US-09-543-681A-7623
22	6	30.0	739	4	US-09-134-001C-3586
23	5	25.0	8	1	US-08-418-594-3
24	5	25.0	8	2	US-08-919-089-3
25	5	25.0	9	4	US-09-311-784A-430
26	5	25.0	12	1	US-08-117-361C-12
27	5	25.0	13	4	US-09-522-666-30

```

28 5 25.0 15 4 US-09-009-953-49 Sequence 49, Appl
29 5 25.0 16 4 US-09-360-237-55 Sequence 55, Appl
30 5 25.0 16 4 US-09-680-571A-83 Sequence 83, Appl
31 5 25.0 16 4 US-09-680-571A-97 Sequence 97, Appl
32 5 25.0 17 1 US-08-487-890A-104 Sequence 104, App
33 5 25.0 17 2 US-08-478-435-104 Sequence 104, App
34 5 25.0 17 2 US-08-337-483-104 Sequence 104, App
35 5 25.0 17 2 US-08-478-373-104 Sequence 104, App
36 5 25.0 17 3 US-08-474-671-104 Sequence 104, App
37 5 25.0 17 3 US-08-483-577A-104 Sequence 104, App
38 5 25.0 17 3 US-08-897-438-104 Sequence 104, App
39 5 25.0 17 4 US-08-637-654-104 Sequence 104, App
40 5 25.0 17 4 US-08-649-518-104 Sequence 104, App
41 5 25.0 22 2 US-09-013-634-8 Sequence 104, App
42 5 25.0 27 4 US-09-360-237-57 Sequence 57, Appl
43 5 25.0 51 4 US-09-621-976-7274 Sequence 7274, Ap
44 5 25.0 53 4 US-09-663-600A-133 Sequence 133, App
45 5 25.0 54 4 US-09-621-976-5883 Sequence 5883, Ap

```

ALIGNMENTS

```

RESULT 1
US-09-886-319A-13
; Sequence 13, Application US/09886319A
; Patent No. 6586185
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; TITLE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014002
; CURRENT APPLICATION NUMBER: US/09/886,319A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 10030149.5
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-886-319A-13

Query Match 35.0%; Score 7; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

Qy 9 FLLWTPY 15
Db 249 FLLWTPY 255

RESULT 2
US-09-673-395A-293
; Sequence 293, Application US/09673395A
; Patent No. 6620923
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE

```

```
; FILE REFERENCE: ALBRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 293
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-673-395A-293

Query Match      30.0%; Score 6; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 VLSFLL 11
DB      4 VLSFLL 9

RESULT 3
US-09-227-357-249
; Sequence 249, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953

; FILE REFERENCE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 249
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (36)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals stop translation
US-09-227-357-249

Query Match      30.0%; Score 6; DB 4; Length 80;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 VLSFLL 11
DB      15 VLSFLL 20

RESULT 4
US-09-482-273-177
; Sequence 177, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 177
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-482-273-177
```

```

Query Match          30.0%; Score 6; DB 4; Length 105;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
   |||||
Db 26 VLSFLL 31

RESULT 5
US-09-328-352-6057
; Sequence 6057, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6057
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6057

Query Match          30.0%; Score 6; DB 4; Length 124;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
   |||||
Db 86 VLSFLL 91

RESULT 6
US-09-489-039A-11733
; Sequence 11733, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 11733
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11733

Query Match          30.0%; Score 6; DB 4; Length 148;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
   |||||
Db 14 VLSFLL 19

RESULT 7
US-09-313-458-34
; Sequence 34, Application US/09313458
; Patent No. 6576743
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Haldeman, Betty A.

```

```

; TITLE OF INVENTION: MAMMALIAN CYTOKINE-LIKE POLYPEPTIDE-10
; FILE REFERENCE: 97-72
; CURRENT APPLICATION NUMBER: US/09/313,458
; CURRENT FILING DATE: 1999-05-17
; EARLIER APPLICATION NUMBER: 09/199,586
; EARLIER FILING DATE: 1998-11-25
; EARLIER APPLICATION NUMBER: 60/066,597
; EARLIER FILING DATE: 1997-11-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-313-458-34

Query Match          30.0%; Score 6; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FLLWTP 14
   |||||
Db 16 FLLWTP 21

RESULT 8
US-09-746-359A-8
; Sequence 8, Application US/09746359A
; Patent No. 6610286
; GENERAL INFORMATION:
; APPLICANT: Thompson, Penny
; APPLICANT: Foster, Donald C.
; APPLICANT: Xu, Wenfeng
; APPLICANT: Madden, Karen L.
; APPLICANT: Kelly, James D.
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Blumberg, Hal
; APPLICANT: Eagan, Maribeth A.
; APPLICANT: Jaspers, Stephen R.
; APPLICANT: Chandrasekhar, Yasmin A.
; APPLICANT: No. 6610286ak, Julia E.
; TITLE OF INVENTION: Method for Treating Inflammation
; FILE REFERENCE: 99-108
; CURRENT APPLICATION NUMBER: US/09/746,359A
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/171,969
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 60/213,341
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-746-359A-8

Query Match          30.0%; Score 6; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FLLWTP 14
   |||||
Db 16 FLLWTP 21

RESULT 9
US-09-313-458-19
; Sequence 19, Application US/09313458
; Patent No. 6576743
; GENERAL INFORMATION:
; APPLICANT: Conklin, Darrell C.
; APPLICANT: Haldeman, Betty A.

```

;/ TITLE OF INVENTION: MAMMALIAN CYTOKINE-LIKE POLYPEPTIDE-10
;/ FILE REFERENCE: 97-72
;/ CURRENT APPLICATION NUMBER: US/09/313,458
;/ CURRENT FILING DATE: 1999-05-17
;/ EARLIER APPLICATION NUMBER: 09/199,586
;/ EARLIER FILING DATE: 1998-11-25
;/ EARLIER APPLICATION NUMBER: 60/066,597
;/ EARLIER FILING DATE: 1997-11-26
;/ NUMBER OF SEQ ID NOS: 43
;/ SOFTWARE: FastSeq for Windows Version 3.0
;/ SEQ ID NO 19
;/ LENGTH: 176
;/ TYPE: PRT
;/ ORGANISM: Mus musculus
US-09-313-458-19

Query Match 30.0%; Score 6; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 FLWTWP 14
Db 16 FLWTWP 21

RESULT 10
US-09-746-359A-5
;/ Sequence 5, Application US/09746359A
;/ Patent No. 6610286
;/ GENERAL INFORMATION:
;/ APPLICANT: Thompson, Penny
;/ APPLICANT: Foster, Donald C.
;/ APPLICANT: Xu, Wenfeng
;/ APPLICANT: Madden, Karen L.
;/ APPLICANT: Kelly, James D.
;/ APPLICANT: Sprecher, Cindy A.
;/ APPLICANT: Blumberg, Hal
;/ APPLICANT: Eagan, Maribeth A.
;/ APPLICANT: Jaspers, Stephen R.
;/ APPLICANT: Chandrasekhar, Yasmin A.
;/ APPLICANT: No. 6610286ak, Julia E.
;/ TITLE OF INVENTION: Method for Treating Inflammation
;/ FILE REFERENCE: 99-108
;/ CURRENT APPLICATION NUMBER: US/09/746,359A
;/ CURRENT FILING DATE: 2001-05-21
;/ PRIOR APPLICATION NUMBER: 60/171,969
;/ PRIOR FILING DATE: 1999-12-23
;/ PRIOR APPLICATION NUMBER: 60/213,341
;/ PRIOR FILING DATE: 2000-06-22
;/ NUMBER OF SEQ ID NOS: 72
;/ SOFTWARE: FastSeq for Windows Version 3.0
;/ SEQ ID NO 5
;/ LENGTH: 176
;/ TYPE: PRT
;/ ORGANISM: Mus musculus
US-09-746-359A-5

Query Match 30.0%; Score 6; DB 4; Length 176;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 FLWTWP 14
Db 16 FLWTWP 21

RESULT 11
US-09-252-991A-18224
;/ Sequence 18224, Application US/09252991A
;/ Patent No. 6551795
;/ GENERAL INFORMATION:
;/ APPLICANT: Marc J. Rubenfield et al.
;/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

;/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;/ FILE REFERENCE: 107196.136
;/ CURRENT APPLICATION NUMBER: US/09/252,991A
;/ CURRENT FILING DATE: 1999-02-18
;/ PRIOR APPLICATION NUMBER: US 60/074,788
;/ PRIOR FILING DATE: 1998-02-18
;/ PRIOR APPLICATION NUMBER: US 60/094,190
;/ PRIOR FILING DATE: 1998-07-27
;/ NUMBER OF SEQ ID NOS: 33142
;/ SEQ ID NO 18224
;/ LENGTH: 215
;/ TYPE: PRT
;/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18224

Query Match 30.0%; Score 6; DB 4; Length 215;
Best Local Similarity 100.0%; Pred. No. 90;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 VLSFLL 11
Db 41 VLSFLL 46

RESULT 12
US-09-107-532A-6781
;/ Sequence 6781, Application US/09107532A
;/ Patent No. 6583275
;/ GENERAL INFORMATION:
;/ APPLICANT: Lynn A Doucette-Stamm and David Bush
;/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;/ ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;/ NUMBER OF SEQUENCES: 7310
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSER: GENOME THERAPEUTICS CORPORATION
;/ STREET: 100 Beaver Street
;/ CITY: Waltham
;/ STATE: Massachusetts
;/ COUNTRY: USA
;/ ZIP: 02154
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: CD-ROM ISO9660
;/ COMPUTER: PC
;/ OPERATING SYSTEM: <Unknown>
;/ SOFTWARE: ASCII
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/09/107,532A
;/ FILING DATE: 30-Jun-1998
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: 60/085,598
;/ FILING DATE: 14 May 1998
;/ APPLICATION NUMBER: 60/051571
;/ FILING DATE: July 2, 1997
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Ariniello, Pamela Deneke
;/ REGISTRATION NUMBER: 40,489
;/ REFERENCE/DOCKET NUMBER: GTC-012
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (781)893-5007
;/ TELEFAX: (781)893-8277
;/ INFORMATION FOR SEQ ID NO: 6781:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 338 amino acids
;/ TYPE: amino acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ HYPOTHETICAL: YES
;/ ORIGINAL SOURCE:
;/ ORGANISM: Enterococcus faecium
;/ FEATURE:
;/ NAME/KEY: misc feature
;/ LOCATION: (B) LOCATION 1...338
;/ SEQUENCE DESCRIPTION: SEQ ID NO: 6781:

US-09-107-532A-6781

Query Match 30.0%; Score 6; DB 4; Length 338;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
| | | | |
Db 302 VLSFLL 307

RESULT 13

US-09-482-273-113
; Sequence 113, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 113
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (53)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (345)
; OTHER INFORMATION: Xaa equals stop translation
US-09-482-273-113

Query Match 30.0%; Score 6; DB 4; Length 345;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
| | | | |
Db 265 VLSFLL 270

RESULT 14

US-09-482-273-208
; Sequence 208, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 208
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (78)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-482-273-208

Query Match 30.0%; Score 6; DB 4; Length 369;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
| | | | |
Db 290 VLSFLL 295

RESULT 15

US-08-153-848-42
; Sequence 42, Application US/08153848
; Patent No. 5759804
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/153,848
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5759804and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-153-848-42

Query Match 30.0%; Score 6; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
| | | | |
Db 168 VLSFLL 173

Search completed: July 19, 2004, 17:15:15
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:14:36 ; Search time 42 Seconds
(without alignments)
148.838 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1285345 seqs, 312560633 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	19	95.0	20	14	US-10-195-730-116
2	7	35.0	57	11	US-09-864-408A-2008
3	7	35.0	93	12	US-10-424-599-165165
4	7	35.0	355	14	US-10-376-564-13
5	7	35.0	1019	9	US-09-738-626-6499
6	7	35.0	1019	15	US-10-380-055-12
7	7	35.0	1369	16	US-10-437-963-184093
8	6	30.0	51	12	US-10-424-599-147607
9	6	30.0	54	12	US-10-424-599-148937
10	6	30.0	52	16	US-10-437-963-174252
11	6	30.0	66	11	US-09-864-408A-7962
12	6	30.0	62	12	US-10-424-599-187640
13	6	30.0	65	12	US-10-424-599-173594
14	6	30.0	68	16	US-10-437-963-146491
15	6	30.0	72	12	US-10-424-599-205684

16	6	30.0	73	15	US-10-242-355-459	Sequence 459, App
17	6	30.0	77	12	US-10-424-599-281026	Sequence 281026,
18	6	30.0	79	12	US-09-973-278-260	Sequence 260, App
19	6	30.0	80	10	US-09-983-802-249	Sequence 249, App
20	6	30.0	80	12	US-09-984-490-249	Sequence 249, App
21	6	30.0	83	9	US-09-796-692-1812	Sequence 1812, App
22	6	30.0	83	9	US-09-796-692-2069	Sequence 2069, App
23	6	30.0	83	14	US-10-040-862-1812	Sequence 1812, App
24	6	30.0	83	14	US-10-040-862-2069	Sequence 2069, App
25	6	30.0	83	15	US-10-057-4758-1812	Sequence 1812, App
26	6	30.0	83	15	US-10-057-4758-2069	Sequence 2069, App
27	6	30.0	83	15	US-10-154-8848-1812	Sequence 1812, App
28	6	30.0	83	15	US-10-154-8848-2069	Sequence 2069, App
29	6	30.0	84	16	US-10-437-963-127297	Sequence 127297,
30	6	30.0	88	12	US-10-424-599-276079	Sequence 276079,
31	6	30.0	92	14	US-10-029-386-33935	Sequence 33935, A
32	6	30.0	99	10	US-09-764-891-3040	Sequence 3040, App
33	6	30.0	105	10	US-09-984-271-177	Sequence 177, App
34	6	30.0	105	12	US-09-984-276-177	Sequence 196176,
35	6	30.0	109	12	US-10-424-599-149648	Sequence 149648,
36	6	30.0	117	12	US-10-424-599-236063	Sequence 236063,
37	6	30.0	119	12	US-10-424-599-236063	Sequence 236063,
38	6	30.0	123	12	US-10-425-114-45129	Sequence 45129, A
39	6	30.0	141	12	US-10-424-599-189984	Sequence 189984,
40	6	30.0	149	12	US-10-425-114-50534	Sequence 50534, A
41	6	30.0	154	9	US-09-746-359A-8	Sequence 8, Appli
42	6	30.0	154	14	US-10-321-163-34	Sequence 34, Appl
43	6	30.0	154	14	US-10-413-661-34	Sequence 34, Appl
44	6	30.0	154	15	US-10-424-658-8	Sequence 8, Appli
45	6	30.0	154	16	US-10-471-151-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1

US-10-195-730-116
; Sequence 116, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 116
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (20)
; OTHER INFORMATION: Xaa equals stop translation
US-10-195-730-116

Query Match 95.0%; Score 19; DB 14; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.8e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAHSVLSFLWTPYALKS 19

Db 1 MAHSVLSFLWTPYALKS 19

RESULT 2

```
US-09-864-408A-2008
; Sequence 2008, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2008
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-2008

Query Match          35.0%; Score 7; DB 11; Length 57;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 SVLSFLL 11
Db      19 SVLSFLL 25

RESULT 3
US-10-424-599-165165
; Sequence 165165, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 165165
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_12015C.1.pap
US-10-424-599-165165

Query Match          35.0%; Score 7; DB 12; Length 93;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 SVLSFLL 11
Db      35 SVLSFLL 41

RESULT 4
US-10-376-564-13
; Sequence 13, Application US/10376564
; Publication No. US20030180302A1
; GENERAL INFORMATION:
; APPLICANT: Wolf, Eckhard
; APPLICANT: Werner, Sabine
; APPLICANT: Halle, Jorn-Peter
; APPLICANT: Regenbogen, Johannes
; APPLICANT: Goppelt, Andreas
; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for
; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound

US-09-864-408A-2008
; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically
; FILE OF INVENTION: Active Substances
; FILE REFERENCE: 50125/014003
; CURRENT APPLICATION NUMBER: US/10/376,564
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 09/886,319
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/222,081
; PRIOR FILING DATE: 2000-08-01
; PRIOR APPLICATION NUMBER: DE 100 30 149.5-41
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-376-564-13

Query Match          35.0%; Score 7; DB 14; Length 355;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 FLWTPY 15
Db      249 FLWTPY 255

RESULT 5
US-09-738-626-6499
; Sequence 6499, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6499
; LENGTH: 1019
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6499

Query Match          35.0%; Score 7; DB 9; Length 1019;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 SVLSFLL 11
Db      141 SVLSFLL 147

RESULT 6
US-10-380-055-12
; Sequence 12, Application US/10380055
```



```
; Publication No. US20040014180A1
; GENERAL INFORMATION:
; APPLICANT: BOTT, Michael, et al.
; TITLE OF INVENTION: A Method For Microbial Production Of Metabolic Products, Polynucle
; TITLE OF INVENTION: From Coryneform Bacteria And Their Use
; FILE REFERENCE: 032301 WN 331
; CURRENT APPLICATION NUMBER: US/10/380,055
; CURRENT FILING DATE: 2003-03-11
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1019
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-380-055-12

Query Match      35.0%; Score 7; DB 15; Length 1019;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 SVLSFLL 11
Db      141 SVLSFLL 147

RESULT 7
US-10-437-963-184093
; Sequence 184093, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 184093
; LENGTH: 1369
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_81119C.1.pep
US-10-437-963-184093

Query Match      35.0%; Score 7; DB 16; Length 1369;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 VLSFLW 12
Db      742 VLSFLW 748

RESULT 8
US-10-424-599-147607
; Sequence 147607, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
```

```
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 147607
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_104307C.1.pep
US-10-424-599-147607

Query Match      30.0%; Score 6; DB 12; Length 51;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 VLSFLL 11
Db      23 VLSFLL 28

RESULT 9
US-10-424-599-148937
; Sequence 148937, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 148937
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_105512C.1.pep
US-10-424-599-148937

Query Match      30.0%; Score 6; DB 12; Length 54;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 SVLSFL 10
Db      18 SVLSFL 23

RESULT 10
US-10-437-963-174252
; Sequence 174252, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 174252
; LENGTH: 56
```

; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_72210C.1.pep
US-10-437-963-174252

Query Match 30.0%; Score 6; DB 16; Length 56;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LSFLW 12
Db 47 LSFLW 52

RESULT 11
US-09-864-408A-7962
; Sequence 7962, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7962
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-7962

Query Match 30.0%; Score 6; DB 11; Length 62;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFL 10
Db 32 SVLSFL 37

RESULT 12
US-10-424-599-187640
; Sequence 187640, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 187640
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(62)
; OTHER INFORMATION: unsure at all xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_140452C.1.pep
US-10-424-599-187640

Query Match 30.0%; Score 6; DB 12; Length 62;

Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
Db 32 VLSFLL 37

RESULT 13
US-10-424-599-173594
; Sequence 173594, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 173594
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_127775C.1.pep
US-10-424-599-173594

Query Match 30.0%; Score 6; DB 12; Length 65;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LSFLW 12
Db 46 LSFLW 51

RESULT 14
US-10-437-963-146491
; Sequence 146491, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 146491
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_4710C.1.pep
US-10-437-963-146491

Query Match 30.0%; Score 6; DB 16; Length 68;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LSFLW 12
Db 46 LSFLW 51

Db 50 LSFLW 55

RESULT 15
US-10-424-599-205684
; Sequence 205684, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION: Thomas J
; APPLICANT: La Rosa
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 205684
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_27760C.1.pep
US-10-424-599-205684

Query Match 30.0%; Score 6; DB 12; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 AHSVLS 8
| | | | |
Db 54 AHSVLS 59

Search completed: July 19, 2004, 17:19:49
Job time : 42 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:11:12 ; Search time 16 Seconds
(without alignments)
120.239 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHVSLSFLLWTPYALKSX 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR_78:*
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	35.0	355	2 I49339	macrophage inflamm
2	7	35.0	1865	2 G86152	T7123.15 protein -
3	6	30.0	57	1 W9BPC7	gene 19.3 protein
4	6	30.0	59	2 T07432	photosystem II pro
5	6	30.0	82	2 T06236	probable 3,4-dihyd
6	6	30.0	87	2 F87485	flhB-related prote
7	6	30.0	108	2 C72852	AcOrf-19 protein -
8	6	30.0	108	2 D44221	orf4 protein - Aut
9	6	30.0	110	2 T41765	AcMNPV orf19 - Bom
10	6	30.0	111	2 F71032	hypothetical prote
11	6	30.0	180	2 H90450	hypothetical prote
12	6	30.0	183	2 B83183	hypothetical prote
13	6	30.0	205	2 F85515	unknown protein en
14	6	30.0	205	2 C90655	hypothetical prote
15	6	30.0	246	2 AH3615	gdp-4-dehydro-D-ri
16	6	30.0	247	1 E69021	conserved hypotet
17	6	30.0	255	2 T25853	hypothetical prote
18	6	30.0	264	2 D71037	hypothetical prote
19	6	30.0	264	2 AD2924	hypothetical prote
20	6	30.0	264	2 C97698	chromosome partiti
21	6	30.0	276	2 F72342	conserved hypotet
22	6	30.0	286	1 RLPUGG	rRNA N-glycosidase
23	6	30.0	292	2 C70421	conserved hypotet
24	6	30.0	297	2 A45335	cytochrome-c oxida
25	6	30.0	297	2 T19888	hypothetical prote
26	6	30.0	297	2 G87363	metallo-beta-lacta
27	6	30.0	318	2 C83209	probable oxidoredu
28	6	30.0	344	2 F81170	tetraacyldisacchar
29	6	30.0	344	2 C81933	probable tetraacyl

30 6 30.0 354 2 DB6843
31 6 30.0 372 2 S72711
32 6 30.0 375 2 E83954
33 6 30.0 386 2 A73133
34 6 30.0 391 2 C84268
35 6 30.0 397 2 A96681
36 6 30.0 399 2 H84087
37 6 30.0 400 2 C82503
38 6 30.0 422 2 C75631
39 6 30.0 430 2 AF1356
40 6 30.0 430 2 AG1726
41 6 30.0 431 2 A72245
42 6 30.0 440 2 C96014
43 6 30.0 453 2 T39240
44 6 30.0 456 2 T40386
45 6 30.0 456 2 H90723

ALIGNMENTS

RESULT 1

I49339
macrophage inflammatory protein-1 alpha receptor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
C:Accession: I49339
R:Gao, J.L.; Murphy, P.M.
J. Biol. Chem. 270, 17494-17501, 1995
A:Title: Cloning and differential tissue-specific expression of three mouse beta chemoki:
A:Reference number: I49339; MUID:95340546; PMID:7542241
A:Accession: I49339
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-355 <RES>
A:Cross-references: EMBL:U28404; NID:g881547; PIDN:AAA89153.1; PID:g881548
C:Superfamily: vertebrate rhodopsin

Query Match 35.0%; Score 7; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 FLLWTPY 15
|||||
Db 249 FLLWTPY 255

RESULT 2

G86152
T7123.15 protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 27-Nov-2001
C:Accession: G86152
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali, I.
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86152
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1865 <STO>
A:Cross-references: GB:AE005172; NID:g3258569; PIDN:AAC24379.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 35.0%; Score 7; DB 2; Length 1865;

Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 HSVLSFL 10
Db 178 HSVLSFL 184

RESULT 3
WBPC7
gene 19.3 protein - phase T7
C:Species: phase T7
C:Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 23-Jul-1999
C:Accession: A04415; S42340
R:Dunn, J.J.; Thompson, K.
submitted to the Nucleic Acid Sequence Database, September 1982
A:Reference number: A94615
A:Accession: A04415
A:Molecule type: DNA
A:Residues: 1-57 <DUN>
R:Dunn, J.J.; Studier, F.W.
J. Mol. Biol. 186, 477-535, 1983
A:Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genes
A:Reference number: S42283; MUID:83241725; PMID:6864790
A:Accession: S42340
A:Molecule type: DNA
A:Residues: 1-57 <DUN>
A:Cross-references: EMBL:V01146; NID:g431187; PIDN:CAA24442.1; PID:g15618
A:Note: the authors did not translate the codon for residue 1
C:Genetics:
A:Gene: 19.3
A:Map position: 96.53-96.96
C:Superfamily: phase T7 gene 19.3 protein

Query Match 30.0%; Score 6; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFL 10
Db 38 SVLSFL 43

RESULT 4
T07432
Photosystem II protein psbX - Japanese black pine chloroplast
C:Species: chloroplast Pinus thunbergiana (Japanese black pine)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 18-Aug-2000
C:Accession: T07432
R:Wakasugi, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugiura, M.
Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994
A:Title: Loss of all ndh genes as determined by sequencing the entire chloroplast genome
A:Reference number: Z16030; MUID:95024047; PMID:7937893
A:Accession: T07432
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-59 <WAK>
A:Cross-references: EMBL:D17510; NID:g529643; PIDN:BAA04312.1; PID:g1262593
C:Genetics:
A:Gene: psbX
A:Genome: chloroplast
C:Superfamily: photosystem II protein psbX
C:Keywords: chloroplast; membrane-associated complex; photosynthesis; photosystem II; th

Query Match 30.0%; Score 6; DB 2; Length 59;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFL 11
Db 43 VLSFL 48

RESULT 5
T06236
probable 3,4-dihydroxy-2-butanone kinase - soybean (fragment)
C:Species: Glycine max (soybean)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 08-Sep-2000
C:Accession: T06236
R:Mahalingam, R.; Knap, H.T.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z15555
A:Accession: T06236
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-82 <MAH>
A:Cross-references: EMBL:AF047054; NID:g2905783; PIDN:AAC03560.1; PID:g2905784
A:Experimental source: cultivar Essex; root
C:Genetics:
A:Gene: dbk
C:Superfamily: glycerone kinase DAK1

Query Match 30.0%; Score 6; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAHSV 7
Db 32 AAHSV 37

RESULT 6
F87485
flhB-related protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: F87485
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hart, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: F87485
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-87 <STO>
A:Cross-references: GB:AE005673; NID:g13423358; PIDN:AAK23882.1; GSPDB:GN00148
C:Genetics:
A:Gene: CCL907

Query Match 30.0%; Score 6; DB 2; Length 87;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFL 11
Db 80 VLSFL 85

RESULT 7
C72852
Acorf-19 protein - Autographa californica nuclear polyhedrosis virus
C:Species: Autographa californica nuclear polyhedrosis virus, ACMPV
A:Note: daDNA virus
C:Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 11-Jan-2000
C:Accession: C72852
R:Ayres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.
Virology 202, 586-605, 1994
A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus
A:Reference number: A72850; MUID:94303173; PMID:8030224
A:Accession: C72852
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-108 <AYR>
A:Cross-references: GB:L22858; NID:g510708; PIDN:AAA66649.1; PID:g559088

C;Genetics:
A;Gene: Acorf-19
C;Superfamily: AcMNPV hypothetical protein 19

Query Match 30.0%; Score 6; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
| | | | |
Db 99 VLSFLL 104

RESULT 8
D44221
orf4 protein - Autographa californica nuclear polyhedrosis virus
C;Species: Autographa californica nuclear polyhedrosis virus, AcMNPV
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 11-Jan-2000
C;Accession: D44221
R;Braunagel, S.C.; Daniel, K.D.; Reilly, L.M.; Guarino, L.A.; Hong, T.; Summers, M.D.
Virology 131, 1003-1008, 1992
A;Title: Sequence, genomic organization of the EcoRI-A fragment of Autographa californica
VP8 of rotavirus.
A;Reference number: A44221; MUID:93079853; PMID:1333113
A;Accession: D44221
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-108 <BRA>
A;Cross-references: GB:S52569
C;Superfamily: AcMNPV hypothetical protein 19

Query Match 30.0%; Score 6; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
| | | | |
Db 99 VLSFLL 104

RESULT 9
T41765
AcMNPV orf19 - Bombyx mori nuclear polyhedrosis virus (isolate T3)
C;Species: Bombyx mori nuclear polyhedrosis virus, BmSNPV
A;Variety: isolate T3
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C;Accession: T41765
R;Gomi, S.; Majima, K.; Maeda, S.
J. Gen. Virol. 80, 1323-1337, 1999
A;Title: Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus.
A;Reference number: 222020; MUID:99281911; PMID:10355780
A;Accession: T41765
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-110 <KAM>
A;Cross-references: EMBL:L33180; NID:g3745835; PIDN:AAC63694.1; PID:g3745847
A;Experimental source: isolate T3
C;Genetics:
A;Note: Orf 11
C;Superfamily: AcMNPV hypothetical protein 19

Query Match 30.0%; Score 6; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
| | | | |
Db 101 VLSFLL 106

RESULT 10
F71032
hypothetical protein PH1550 - Pyrococcus horikoshii

C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
C;Accession: F71032
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin-
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a-
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: F71032
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-111 <KAW>
A;Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30662.1; PID:g3257979
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1550
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1550

Query Match 30.0%; Score 6; DB 2; Length 111;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 LSFLW 12
| | | | |
Db 103 LSFLW 108

RESULT 11
H90450
hypothetical protein SSO2753 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: H90450
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayer, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: H90450
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-180 <KUR>
A;Cross-references: GB:AE006641; NID:gl3816084; PIDN:AAK42863.1; GSPDB:GN00155
C;Genetics:
A;Gene: SSO2753

Query Match 30.0%; Score 6; DB 2; Length 180;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFLL 11
| | | | |
Db 73 VLSFLL 78

RESULT 12
B83183
hypothetical protein PA3698 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: B83183
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bri-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho-
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: B83183
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-183 <STO>

A;Cross-references: GB:AE004789; GB:AE004091; NID:g9949857; PIDN:AAG07086.1; GSPDB:GN001
A;Experimental source: strain PA01
C;Genetics:
A;Gene: FA3698

Query Match 30.0%; Score 6; DB 2; Length 183;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFL 11
|||
Db 9 VLSFL 14

RESULT 13
F85515
unknown protein encoded in prophage CP-933I [imported] - Escherichia coli (strain O157:H7)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: F85515

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamocousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: F85515
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-205 <STO>
A;Cross-references: GB:AE005174; NID:gl2513038; PIDN:AAG54586.1; GSPDB:GN00145; UWGP:Z03
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z0326

Query Match 30.0%; Score 6; DB 2; Length 205;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AHSVLS 8
|||
Db 174 AHSVLS 179

RESULT 14
C90665
hypothetical protein ECs0291 [imported] - Escherichia coli (strain O157:H7, substrain R1
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: C90665

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90665
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-205 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAE33714.1; PID:gl33359748; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain R1MD 0509952
C;Genetics:
A;Gene: ECs0291

Query Match 30.0%; Score 6; DB 2; Length 205;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AHSVLS 8
|||
Db 174 AHSVLS 179

RESULT 15

AH3615
gdp-4-dehydro-D-rhamnose reductase (EC 1.1.1.187) [imported] - Brucella melitensis (stra
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C;Accession: AH3615
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,
; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A;Reference number: AD3252; PMID:11756688
A;Accession: AH3615
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-246 <KUR>
A;Cross-references: GB:AE008918; PIDN:AAL54091.1; PID:gl7985049; GSPDB:GN00191
A;Experimental source: strain 16M

C;Genetics:
A;Gene: EMEI10849
A;Map position: II
C;Keywords: oxidoreductase

Query Match 30.0%; Score 6; DB 2; Length 246;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAHSVLS 7
|||
Db 47 AAHSVLS 52

Search completed: July 19, 2004, 17:14:46
Job time: 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:07:32 ; Search time 13 Seconds
(without alignments)
80.108 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHSLVSLFLWTPYALKSX 20

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	35.0	355	1 CKR1_MOUSE	P51675 mus musculus
2	6	30.0	56	1 PSBK_PINTH	P41598 pinus thunb
3	6	30.0	57	1 V193_BPT7	P03790 bacterioph
4	6	30.0	82	1 DHBK_SOYEN	O49227 glycine max
5	6	30.0	92	1 YB41_STRMU	P59154 streptococ
6	6	30.0	108	1 Y019_NPVAC	P41424 autographa
7	6	30.0	156	1 ARQ1_RALSO	Q8XVP5 ralstonia s
8	6	30.0	176	1 IL20_MOUSE	Q9JKV9 mus musculu
9	6	30.0	285	1 PSD_VIBPA	Q87K29 vibrio para
10	6	30.0	286	1 RIPI_MOMCH	P16094 momordica c
11	6	30.0	297	1 CTA4_BACPF	Q04443 bacillus ps
12	6	30.0	317	1 O6N2_HUMAN	Q8NGY6 homo sapien
13	6	30.0	319	1 O6K6_HUMAN	Q8NGW6 homo sapien
14	6	30.0	322	1 O5T3_HUMAN	Q8NGG3 homo sapien
15	6	30.0	330	1 OR08_MOUSE	Q8VF13 mus musculu
16	6	30.0	344	1 LPXK_NEIMA	Q9JVE4 neisseria m
17	6	30.0	344	1 LPXK_NEIMB	Q9K0D7 neisseria m
18	6	30.0	344	1 ZDH4_HUMAN	Q9NP98 homo sapien
19	6	30.0	352	1 L4R1_HUMAN	Q15722 homo sapien
20	6	30.0	377	1 OPS2_HEMSA	Q25158 hemigrapsus
21	6	30.0	456	1 ENTI_RAT	Q54698 rattus norv
22	6	30.0	458	1 TCR_BACST	P07561 bacillus st
23	6	30.0	458	1 TCR_STRAG	P36890 streptococ
24	6	30.0	458	1 TCR_STRAG	P13924 streptococ
25	6	30.0	458	1 TCR_STRPN	P11063 streptococ
26	6	30.0	461	1 PEX2_PICPA	Q01964 picchia past
27	6	30.0	477	1 DTAL_RHILO	Q98AV2 rhizobium l
28	6	30.0	483	1 CTDB_HUMAN	Q9H579 homo sapien
29	6	30.0	492	1 YCLF_BACSU	P94408 bacillus su
30	6	30.0	494	1 KCF1_HUMAN	Q9H3M0 homo sapien
31	6	30.0	513	1 AMY3_SCHPO	O14154 schizosacch
32	6	30.0	552	1 YGXB_BACSU	P37874 bacillus su
33	6	30.0	565	1 MOT8_MOUSE	O70324 mus musculu

RESULT 1

ID	CKR1_MOUSE	STANDARD;	PRT;	355 AA.
AC	P51675; Q91VP9;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	C-C chemokine receptor type 1 (C-C CKR-1) (CCR-1) (CCR1)			
DE	(Macrophage inflammatory protein-1 alpha receptor) (MIP-1alpha-R)			
DE	(RANTES-R).			
GN	CCR1 OR CMKBR1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/Sv; TISSUE=Peritoneal macrophage;			
RX	MEDLINE=96072806; PubMed=7594543;			
RA	Post T.W., Bozic C.R., Rothenberg M.E., Luster A.D., Gerard N.,			
RA	Gerard C.;			
RT	"Molecular characterization of two murine eosinophil beta chemokine			
RT	receptors."			
RL	J. Immunol. 155:5299-5305 (1995).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129/SvJ;			
RX	MEDLINE=95340546; PubMed=7542241;			
RA	Gao J.-L., Murphy P.M.;			
RT	"Cloning and differential			
RT	tissue-specific expression of three mouse			
RT	beta chemokine receptor-like genes, including the gene for a			
RT	functional macrophage inflammatory protein-1 alpha receptor."			
RL	J. Biol. Chem. 270:17494-17501 (1995).			
RL	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Breast;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Reingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,			
RA	Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,			
RA	Boak S.A., McWay P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shervencko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,			
RA	Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length			
RT	human and mouse cDNA sequences."			

ALIGNMENTS

34	6	30.0	579	1	FD21_MYCLE
35	6	30.0	613	1	MOT8_HUMAN
36	6	30.0	649	1	QOX1_BACSU
37	6	30.0	790	1	LY14_YEAST
38	6	30.0	859	1	CHS1_SCHPO
39	6	30.0	1324	1	SMC4_SCHPO
40	6	30.0	1416	1	BLM_MOUSE
41	6	30.0	1556	1	PRO5_DROVI
42	6	30.0	1620	1	CO3_EPTBU
43	6	30.0	2103	1	RRPL_UUK
44	6	30.0	3859	1	TRAP_HUMAN
45	5	25.0	56	1	NU6M_ARTSA

P54200	mycobacteri
P36021	homo sapien
P34956	bacillus su
P40971	saccharomyc
P30597	schizosacch
P41004	schizosacch
O88700	mus musculu
Q9u6a1	drosophila
P98094	eptaretus
P33453	ukuniemi v
Q9y4a5	homo sapien
P19048	artemia sal

```

Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
-!- FUNCTION: RECEPTOR FOR A C-C TYPE CHEMOKINE. BINDS TO MIP-1-ALPHA,
RANTES, AND LESS EFFICIENTLY, TO MIP-1-BETA OR MCP-1 AND
SUBSEQUENTLY TRANSDUCES A SIGNAL BY INCREASING THE INTRACELLULAR
CALCIUM IONS LEVEL. RESPONSIBLE FOR AFFECTING STEM CELL
PROLIFERATION.
CC
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC
CC -!- TISSUE SPECIFICITY: Detected in the heart, spleen, lung,
CC peritoneal exudate cells and leukocytes.
CC
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U29678; AA86119.1; -.
CC EMBL; U28404; AA89153.1; -.
CC EMBL; BC011092; AAH11092.1; -.
CC F1R; I49339; I49339.
CC MGD; MGI:104618; Ccrl.
CC GO; GO:0016493; F:C-C chemokine receptor activity; IDA.
CC GO; GO:0005515; F:protein binding; IPI.
CC GO; GO:0030595; P:immune cell chemotaxis; IDA.
CC GO; GO:0006954; P:inflammatory response; IMP.
CC GO; GO:0030099; P:myeloid blood cell differentiation; IMP.
CC InterPro; IPR000276; GPCR_Rhodopsin.
CC Pfam; PF00001; 7tm 1; 1.
CC PRINTS; PR00237; GPCR_RHODOPSIN.
CC PROSITE; PS00237; G PROTEIN RECP F1_1; 1.
CC PROSITE; PS0262; G PROTEIN RECP F1_2; 1.
CC PROSITE; PS0262; G PROTEIN RECP F1_2; 1.
KW G-protein coupled receptor; Transmembrane.
FT DOMAIN 1 34 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 35 60 1 (POTENTIAL).
FT DOMAIN 61 64 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 65 91 2 (POTENTIAL).
FT DOMAIN 92 107 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 108 129 3 (POTENTIAL).
FT DOMAIN 130 146 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 147 171 4 (POTENTIAL).
FT DOMAIN 172 197 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 198 223 5 (POTENTIAL).
FT DOMAIN 224 239 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 240 264 6 (POTENTIAL).
FT DOMAIN 265 281 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 282 305 7 (POTENTIAL).
FT DOMAIN 306 355 CYTOPLASMIC (POTENTIAL).
FT DISULFID 106 183 BY SIMILARITY.
FT CONFLICT 55 55 M -> V (IN REF. 2 AND 3).
FT CONFLICT 149 149 L -> P (IN REF. 3).
FT CONFLICT 278 278 H -> Q (IN REF. 3).
SQ SEQUENCE 355 AA; 40901 MW; FCE9FF70E6F38B1 CRC64;

Query Match 35.0%; Score 7; DB 1; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FILWTPY 15
Db 249 FILWTPY 255

RESULT 2
ID PSBK_PINTH STANDARD; PRT; 56 AA.
AC P41598;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Photosystem II reaction center protein K precursor (PSII-K).

Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
-!- FUNCTION: This protein is a component of the reaction center of
Photosystem II.
CC
CC -!- SIMILARITY: Belongs to the psbK family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; DJ7510; BAA04312.1; ALT_INIT.
CC HAMAP; MF_00441; -. 1.
CC InterPro; IPR003687; PSII_PsbK.
CC Pfam; PF02533; PsbK; 1.
CC Photosystem II; Chloroplast.
FT PROPEP 1 19 POTENTIAL.
FT CHAIN 20 56 PHOTOSYSTEM II REACTION CENTER PROTEIN K.
SQ SEQUENCE 56 AA; 6346 MW; 18B1DEEF198ACASA CRC64;

Query Match 30.0%; Score 6; DB 1; Length 56;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFLL 11
Db 40 VLSFLL 45

RESULT 3
ID V193_BPT7 STANDARD; PRT; 57 AA.
AC P03790;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-1989 (Rel. 10, Last annotation update)
DE Gene 19.3 protein.
GN 19.3.
OS Bacteriophage T7.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=10760;
RN [1]_TaxID=10760;
RP SEQUENCE FROM N.A.
RX MEDLINE=83241725; PubMed=6864790;
RA Dunn J.J., Studier F.W.;
RT "Complete nucleotide sequence of bacteriophage T7 DNA and the
RL J. Mol. Biol. 166:477-535(1983).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```

```

DR EMBL; V01146; CAA24442.1; -.
DR PIR; A04415; W9BPC7.
SQ SEQUENCE 57 AA; 6561 MW; A16EA477F5203860 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFL 10
Db 38 SVLSFL 43

RESULT 4
ID DBK_SOYEN STANDARD; PRT; 82 AA.
AC O49227;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Putative 3,4-dihydroxy-2-butanone kinase (EC 2.7.1.-) (Fragment).
GN DBK.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OC NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Essex; TISSUE=Root;
RA Herz S., Eberhardt S., Bacher A.;
RT "L. esculentum mRNA for 3,4-dihydroxy-2-butanone kinase.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the dihydroxyacetone kinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A014951; AAN5833.1; -.
DR HAMAP; MF 01041; -.
DR InterPro: IPR007920; UPF0223.
DR Pfam; PF05256; UPF0223, 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 92 AA; 10622 MW; BBC44A96FEF0BD1 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 92;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFL 10
Db 20 SVLSFL 25

RESULT 6
ID Y019 NPVAC STANDARD; PRT; 108 AA.
AC P41424;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 12.2 kDa protein in EGT-1A1 intergenic region (ORF4).
OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OC NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C6;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus.";
RL Virology 202:586-605(1994).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=E2;
CC MEDLINE=93079853; PubMed=1333113;
CC Braunagel S.C., Daniel K.D., Reilly L.M., Guarino L.A., Hong T.,
CC Summers M.D.;
CC "Sequence, genomic organization of the ECoRI-A fragment of Autographa
CC californica nuclear polyhedrosis virus, and identification of a
CC viral-encoded protein resembling the outer capsid protein VP8 of
CC rotavirus.";
CC Virology 191:1003-1008(1992).
CC -1- SIMILARITY: TO CORRESPONDING ORF IN OPMNPV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)

CC EMBL; L22858; AAA66649.1; -
CC EMBL; M96361; AAA66789.1; -
CC PIR; C72852; C72852.
CC PIR; D44221; D44221.
KW Hypothetical protein.
FT CONFLICT 6 A -> R (IN REF. 2).
SQ SEQUENCE 108 AA; 12162 MW; 8363B15DF72486A1 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 VLSFL 11
Db 99 VLSFL 104
|||||

RESULT 7

ARQ1 RALSO
ID ARQ1 RALSO STANDARD; PRT; 156 AA.
AC Q8XVP5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 3-dehydroquininate dehydratase 1 (EC 4.2.1.10) (3-dehydroquinase 1)
(Type II DHQase 1)
GN ARQ01 OR RSC2785 OR R500062.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).

CC -1- FUNCTION: Catalyzes a trans-dehydration via an enolate
intermediate (By similarity).
CC -1- CATALYTIC ACTIVITY: 3-dehydroquininate = 3-dehydroshikimate + H(2)O.
CC -1- PATHWAY: Aromatic amino acids biosynthesis; shikimate pathway;
third step.
CC -1- SIMILARITY: Belongs to the type-II 3-dehydroquinase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)

CC EMBL; AL646071; CAD16492.1; -
CC HAMAP; MF_00169; -; 1
CC InterPro; IPR001874; DHQuinase_II.
CC Pfam; PF01220; DHQuinase_II; 1.
CC ProDom; PD004527; DHQuinase_II; 1.
CC TIGRfams; TIGR01086; aroQ; 1.
CC PROSITE; PS01029; DEHYDROQUINASE_II; 1.
KW Aromatic amino acid biosynthesis; Lyase; Complete proteome.
SQ SEQUENCE 156 AA; 16521 MW; B6BC38F65A94D032 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAHSV 7
Db 9 AAHSV 14
|||||

RESULT 8

IL20_MOUSE
ID IL20_MOUSE STANDARD; PRT; 176 AA.
AC Q9JKV9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Interleukin-20 precursor (IL-20) (Four alpha helix cytokine ZCYTO10).
GN IL20.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skin;
RX MEDLINE=21097717; PubMed=11163236;
RA Blumberg H., Conklin D., Xu W.F., Grossmann A., Brender T.,
RA Carollo S., Eagan M., Foster D., Haldeman B.A., Hammond A., Haugen H.,
RA Jelinek L., Kelly J.D., Madden K., Maurer M.F., Parrish-Novak J.,
RA Prunkard D., Saxon S., Spracher C., Waggle K., West J.,
RA Whitmore T.E., Yao L., Kuechle M.K., Dale B.A., Chandrasekhar Y.A.;
RT "Interleukin 20: discovery, receptor identification, and role in
epidermal function";
RL Cell 104:9-19(2001).

CC -1- FUNCTION: Cytokine that may be involved in epidermal function and
psoriasis. Acts through STAT3.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the IL-10 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch.)

CC EMBL; AF224267; AAF36680.1; -
CC MGD; MGI:1890473; IL20.
CC InterPro; IPR000098; Interleukin_10.
CC ProDom; PD003687; Interleukin_10; 1.
CC SMART; SM00188; IL10; 1.
CC PROSITE; PS00520; INTERLEUKIN_10; 1.
KW Cytokine; Signal.
CC SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 176 INTERLEUKIN-20.
FT DISULFID 33 126 POTENTIAL.
FT DISULFID 80 132 POTENTIAL.
FT DISULFID 81 134 POTENTIAL.
SQ SEQUENCE 176 AA; 20097 MW; 08577AF656574771 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLLWTP 14
Db 16 FLLWTP 21
|||||

RESULT 9

PSD_VIBPA
ID PSD_VIBPA STANDARD; PRT; 285 AA.

AC Q87KZ9;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Phosphatidylserine decarboxylase proenzyme (EC 4.1.1.65) [Contains:
 DE Phosphatidylserine decarboxylase alpha chain; Phosphatidylserine
 DE decarboxylase beta chain].
 DE PSD OR VP2825.
 OS Vibrio parahaemolyticus.
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 CC Vibrionaceae; Vibrio.
 OX NCBI_TaxID=670;
 RN [1]
 RP STRAIN=RIND 2210633 / Serotype O3:K6;
 RC MEDLINE=22508454; PubMed=12620739;
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
 RT "Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism
 RT distinct from that of *V. cholerae*.";
 RL Lancet 361:743-749(2003).
 CC -!- CATALYTIC ACTIVITY: Phosphatidyl-L-serine =
 CC phosphatidylethanolamine + CO(2).
 CC -!- COFACTOR: Pyruvoyl group (By similarity).
 CC -!- SIMILARITY: Belongs to the phosphatidylserine decarboxylase
 CC family. Subfamily 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AP005082; BAC61088.1; -;
 DR HAMAP; MF 00662; -; 1.
 DR InterPro; IPR003817; PS_Dcarboxylase.
 DR Pfam; PF02666; PS_Dcarboxylase; 1.
 KW Phospholipid biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen;
 KW Complete proteome.
 FT CHAIN 1 251 PHOSPHATIDYLSELINE DECARBOXYLASE BETA
 FT CHAIN (BY SIMILARITY).
 FT CHAIN 252 285 PHOSPHATIDYLSELINE DECARBOXYLASE ALPHA
 FT CHAIN (BY SIMILARITY).
 FT SITE 251 252 CLEAVAGE (NONHYDROLYTIC) (BY SIMILARITY).
 FT MOD_RES 252 252 CONVERTED TO A PYRUVOYL GROUP (BY
 FT SIMILARITY).
 FT SEQUENCE 285 AA; 31730 MW; 9CDDC934BFAB4EE2 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 285;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 13 TPYALK 18
 |||||
 Db 279 TPYALK 284
 RESULT 10
 RIPI MOMCH STANDARD; PRT; 286 AA.
 AC P16094; P24697;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Ribosome-inactivating protein momordin I precursor (EC 3.2.2.22) (rRNA
 DE N-glycosidase) (Alpha-momorcharin) (Alpha-MMC).
 OS Momordica charantia (Bitter melon) (Balsam pear).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
 CC eurosids I; Cucurbitales; Cucurbitaceae; Momordica.

OX NCBI_TaxID=3673;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Seed;
 RA MEDLINE=91159486; PubMed=2001404;
 RX Ho W.K.K., Liu S.C., Shaw P.C., Yeung H.W., Ng T.B., Chan W.Y.;
 RT "Cloning of the cDNA of alpha-momorcharin: a ribosome inactivating
 RT protein.";
 RL Biochim. Biophys. Acta 1088:311-314(1991).
 RN [2]
 RP SEQUENCE OF 24-38.
 RC TISSUE=Seed;
 RX MEDLINE=89326691; PubMed=2753596;
 RA Montecucchi P.C., Lazzarini A.M., Barbieri L., Stirpe F., Soria M.,
 RA Lippi D.;
 RT "N-terminal sequence of some ribosome-inactivating proteins.";
 RL Int. J. Pept. Protein Res. 33:263-267(1989).
 RN [3]
 RP SEQUENCE OF 24-70.
 RC TISSUE=Seed;
 RX MEDLINE=89005108; PubMed=3262509;
 RA Casellas P., Dussosoy D., Falasca A.I., Barbieri L.,
 RA Guillemot J.C., Ferrara P., Bolognesi A., Cennini P., Stirpe F.;
 RT Trichokirin, a ribosome-inactivating protein from the seeds of
 RT Trichosanthes kirilowii Maximowicz. Purification, partial
 RT characterization and use for preparation of immunotoxins.";
 RL Eur. J. Biochem. 176:581-588(1988).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=94356447; PubMed=8075985;
 RA Ren J., Wang Y., Dong Y., Stuart D.I.;
 RT "The N-glycosidase mechanism of ribosome-inactivating proteins
 RT implied by crystal structures of alpha-momorcharin.";
 RL Structure 2:7-16(1994).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.16 ANGSTROMS).
 RX MEDLINE=94192822; PubMed=8143869;
 RA Husain J., Tickle I.J., Wood S.P.;
 RT "Crystal structure of momordin, a type I ribosome inactivating
 RT protein from the seeds of Momordica charantia.";
 RL FEBS Lett. 342:154-158(1994).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=95344383; PubMed=7619070;
 RA Huang Q., Liu S., Tang Y., Jin S., Wang Y.;
 RT "Studies on crystal structures, active-centre geometry and
 RT depurinating mechanism of two ribosome-inactivating proteins.";
 RL Biochem. J. 309:285-298(1995).
 CC -!- CATALYTIC ACTIVITY: Endohydrolysis of the N-glycosidic bond at one
 CC specific adenosine on the 28S rRNA.
 CC -!- SIMILARITY: Belongs to the ribosome-inactivating protein family.
 CC Type 1 RIP subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X57682; CAA0869.1; -;
 DR PIR; S14273; RLPUGG.
 DR PDB; 1A8A; 22-JUN-94.
 DR PDB; 1AHB; 22-JUN-94.
 DR PDB; 1AHC; 22-JUN-94.
 DR PDB; 1MOM; 31-MAY-94.
 DR PDB; 1MRG; 07-FEB-95.
 DR PDB; 1MRH; 07-FEB-95.
 DR PDB; 1MRI; 07-FEB-95.
 DR PDB; 1F8Q; 21-NOV-01.
 DR GlycoSuiteDB; P16094; -;
 DR InterPro; IPR001574; RIP.

DR Pfam; PF00161; RIP; 1.
 DR PRINTS; PR00396; SHIGARICIN.
 DR PROSITE; PS00275; SHIGA_RICIN; 1. inhibitor; Hydrolase; Toxin; Signal;
 KW Plant defense; Protein synthesis inhibitor; Hydrolase; Toxin; Signal;
 FT Glycoprotein; 3D-structure.
 FT SIGNAL 1 23
 FT CHAIN 24 269 RIBOSOME-INACTIVATING PROTEIN MOMORDIN I.
 FT PROPP 270 286 MISSING IN MATURE PROTEIN.
 FT ACT_SITE 183 183
 FT CARBOHYD 250 250 N-LINKED (GLNAC. . .).
 FT /FTid=CAR_000082.

FT STRAND 25 28
 FT HELIX 34 47
 FT STRAND 50 54
 FT TURN 55 56
 FT STRAND 57 60
 FT HELIX 66 69
 FT STRAND 70 76
 FT TURN 78 79
 FT STRAND 82 88
 FT TURN 89 92
 FT STRAND 93 99
 FT TURN 100 101
 FT STRAND 102 105
 FT HELIX 109 114
 FT TURN 115 117
 FT TURN 120 121
 FT STRAND 124 127
 FT HELIX 134 141
 FT HELIX 145 147
 FT STRAND 150 150
 FT STRAND 152 162
 FT TURN 163 163
 FT HELIX 167 186
 FT STRAND 187 187
 FT HELIX 188 195
 FT TURN 196 197
 FT STRAND 202 202
 FT HELIX 206 225
 FT TURN 226 230
 FT STRAND 231 238
 FT TURN 240 241
 FT STRAND 246 250
 FT TURN 251 252
 FT HELIX 254 258
 FT TURN 259 259
 FT STRAND 260 260
 FT STRAND 263 263
 FT HELIX 266 268
 SQ SEQUENCE 286 AA; 31532 MW; E1B013ABEBC216CF CRC64;

Query Match 30.0%; Score 6; DB 1; Length 286;
 Best Local Similarity 100.0%; Pred.No.24;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFL 10
 Db 5 SVLSFL 10

RESULT 11
 ID CTAA_BACPF STANDARD; PRT; 297 AA.
 AC Q0443;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Cytochrome A33 controlling protein.
 GN CTAA.
 OS Bacillus pseudofirmus.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OC NCBI_TaxID=79885;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=OF4;
 RX MEDLINE=93107080; PubMed=7678007;
 RA Quirk P.G., Hicks D.B., Krulwich T.A.;
 RT "Cloning of the cta operon from alkaliphilic Bacillus firmus OF4 and
 RT characterization of the pH-regulated cytochrome caa3 oxidase it
 RT encodes.";
 RL J. Biol. Chem. 268:678-685(1993).
 CC -!- FUNCTION: CTA3 IS REQUIRED FOR CYTOCHROME A33 BIOSYNTHESIS AND
 CC INDEPENDENTLY FOR SPOULATION.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M94110; AAA22362.1; -
 DR PIR; A45335; A45335.
 DR InterPro; IPR003780; COX15_Ctaa.
 DR Pfam; PF02628; COX15-Ctaa; 1.
 KW Sporulation; Membrane.
 SQ SEQUENCE 297 AA; 32326 MW; 1488B374029C0E59 CRC64;
 Query Match 30.0%; Score 6; DB 1; Length 297;
 Best Local Similarity 100.0%; Pred.No.25;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 VLSFLL 11
 Db 245 VLSFLL 250
 RESULT 12
 ID O6N2 HUMAN STANDARD; PRT; 317 AA.
 AC Q8NGY6;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Olfactory receptor 6N2.
 GN OR6N2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Suwa M., Sato T., Okouchi I., Arita M., Putami K., Matsumoto S.,
 RA Tautumi S., Aburatani H., Asai K., Akiyama Y.;
 RT "Genome-wide discovery and analysis of human seven transmembrane helix
 RT receptor genes.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Putative odorant receptor.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC -!- DATABASE: NAME=Human Olfactory Receptor Data Exploratorium (HORDE);
 CC WWW="http://bioinformatics.weizmann.ac.il/cgi-bin/HORDE/showgene.pl?key=symbol&id=O6N2";
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB065630; BAC05856.1; -
 DR Genew; HGNC:15035; OR6N2.
 DR InterPro; IPR000276; GPCR_Rhodpen.

```
DR Pfam: PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 25 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 26 46
FT DOMAIN 47 54 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 55 75
FT DOMAIN 76 99 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 100 120
FT DOMAIN 121 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 160
FT DOMAIN 161 197 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 198 217
FT DOMAIN 218 237 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 238 258
FT DOMAIN 259 271 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 272 292
FT DOMAIN 293 317 CYTOPLASMIC (POTENTIAL).
FT DISULFID 97 189 BY SIMILARITY.
FT CARBOHYD 5 N-LINKED (GLCNAC..) (POTENTIAL).
SQ SEQUENCE 317 AA; 35691 MW; D6CAD3D0B0C981D CRC64;

Query Match 30.0%; Score 6; DB 1; Length 317;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SVLSFL 10
Db 64 SVLSFL 69
|||||

RESULT 13
ID O6K6 HUMAN STANDARD; PRT; 319 AA.
AC Q8NGW6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Olfactory receptor 6K6.
GN OR6K6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Suwa M., Sato T., Okouchi I., Arita M., Futami K., Matsumoto S.,
RA Tautumi S., Aburatani H., Asai K., Akiyama Y.;
RT "Genome-wide discovery and analysis of human seven transmembrane helix
RT receptor genes.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Putative odorant receptor.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC -!- DATABASE: NAME=Human Olfactory Receptor Data Explorer (HORDE);
WWW="http://bioinformatics.weizmann.ac.il/cgi-bin/HORDE/showGene.pl?key=symbols&
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB065650; BAC05876.1; -.
Genew; HGNC:15033; OR6K6.
InterPro; IPR000276; GPCR_Rhodopsn.
Pfam; PF00001; 7tm.1; 1.
PRINTS; PR00237; GPCRHHODPSN.
```

```
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS00262; G_PROTEIN_RECEP_F1_2; 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 29 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 30 50
FT DOMAIN 51 58 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 59 79
FT DOMAIN 80 103 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 104 124
FT DOMAIN 125 143 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 144 164
FT DOMAIN 165 200 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 201 220
FT DOMAIN 221 240 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 241 261
FT DOMAIN 262 274 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 275 295
FT DOMAIN 296 319 CYTOPLASMIC (POTENTIAL).
FT DISULFID 101 193 BY SIMILARITY.
FT CARBOHYD 9 N-LINKED (GLCNAC..) (POTENTIAL).
SQ SEQUENCE 319 AA; 35531 MW; 83A9C35DAAA3B4B9 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 319;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SVLSFL 10
Db 68 SVLSFL 73
|||||

RESULT 14
ID O5T3 HUMAN STANDARD; PRT; 322 AA.
AC Q8NGG3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Olfactory receptor 5T3.
GN OR5T3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Suwa M., Sato T., Okouchi I., Arita M., Futami K., Matsumoto S.,
RA Tautumi S., Aburatani H., Asai K., Akiyama Y.;
RT "Genome-wide discovery and analysis of human seven transmembrane helix
RT receptor genes.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Putative odorant receptor. Could also be involved in
CC taste perception.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC -!- DATABASE: NAME=Human Olfactory Receptor Data Explorer (HORDE);
WWW="http://bioinformatics.weizmann.ac.il/cgi-bin/HORDE/showGene.pl?key=symbols&
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AB065837; BAC06056.1; ALT_INIT.
Genew; HGNC:15297; OR5T3.
InterPro; IPR000276; GPCR_Rhodopsn.
Pfam; PF00001; 7tm.1; 1.
PRINTS; PR00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; FALSE_NEG.
```



```

DR PROSITE; PS0262; G PROTEIN RECP F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 37 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 38 58 1 (POTENTIAL).
FT DOMAIN 59 66 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 67 87 2 (POTENTIAL).
FT DOMAIN 88 111 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 112 132 3 (POTENTIAL).
FT DOMAIN 133 151 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 152 172 4 (POTENTIAL).
FT DOMAIN 173 208 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 209 229 5 (POTENTIAL).
FT DOMAIN 230 249 6 (POTENTIAL).
FT TRANSMEM 250 270 7 (POTENTIAL).
FT DOMAIN 271 283 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 284 304 8 (POTENTIAL).
FT DOMAIN 305 322 CYTOPLASMIC (POTENTIAL).
FT DISULFID 109 201 BY SIMILARITY.
FT CARBOHYD 101 17 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 322 AA; 36240 MW; 3BEC6146C34EC6E3 CRC64;

Query Match 30.0%; Score 6; DB 1; Length 322;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFL 10
DB 76 SVLSFL 81

RESULT 15
OR08_MOUSE STANDARD; PRT; 330 AA.
AC QEVFL3;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative olfactory receptor 179-7.
GN MOR179-7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN [2]
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21676863; PubMed=11802173;
RA Zhang X., Firestein S.;
RT "The olfactory receptor gene superfamily of the mouse.";
RL Nat. Neurosci. 5:124-133(2002).
RN [1]
RN [2]
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21864088; PubMed=11875048;
RA Young J.M., Friedman C., Williams E.M., Ross J.A., Tonnes-Priddy L.,
RA Trask B.J.;
RT "Different evolutionary processes shaped the mouse and human
RT olfactory receptor gene families.";
RL Hum. Mol. Genet. 11:535-546(2002).
RN [1]
RN [2]
RN [3]
RP ERRATUM.
RA Young J.M., Friedman C., Williams E.M., Ross J.A., Tonnes-Priddy L.,
RA Trask B.J.;
RL Hum. Mol. Genet. 11:1683-1683(2002).
CC -!- FUNCTION: Potential odorant receptor.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY073728; AAL61391.1; -.
DR EMBL; AY318291; AAP71520.1; -.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCR_Rhodopsn.
DR PROSITE; PS00237; G PROTEIN RECP F1_1; 1.
DR PROSITE; PS0262; G PROTEIN RECP F1_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Olfaction.
FT DOMAIN 1 37 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 38 58 1 (POTENTIAL).
FT DOMAIN 59 66 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 67 87 2 (POTENTIAL).
FT DOMAIN 88 111 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 112 132 3 (POTENTIAL).
FT DOMAIN 133 145 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 146 166 4 (POTENTIAL).
FT DOMAIN 167 208 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 209 229 5 (POTENTIAL).
FT DOMAIN 230 249 6 (POTENTIAL).
FT TRANSMEM 250 270 7 (POTENTIAL).
FT DOMAIN 271 283 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 284 304 7 (POTENTIAL).
FT DOMAIN 305 330 CYTOPLASMIC (POTENTIAL).
FT DISULFID 109 201 BY SIMILARITY.
FT CARBOHYD 17 17 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 330 AA; 37363 MW; 64916C0E6013B87C CRC64;

Query Match 30.0%; Score 6; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFL 10
DB 76 SVLSFL 81

Search completed: July 19, 2004, 17:13:37
Job time : 19 secs

```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2004, 17:10:42 ; Search time 36 Seconds
(without alignments)
175.288 Million cell updates/sec

Title: US-10-799-747-116

Perfect score: 20

Sequence: 1 MAHSVLSFLWTFYALKSX 20

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- SPTREMBL 25:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mbc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriaph:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	35.0	355	11 Q9JLY8	Q9jly8 rattus norv
2	7	35.0	355	11 Q8BYW4	Q8byw4 mus musculu
3	7	35.0	355	11 Q8BMH9	Q8bmh9 mus musculu
4	7	35.0	468	17 Q8TUV5	Q8tuv5 methanopyru
5	7	35.0	525	3 Q74247	Q74247 pichia past
6	7	35.0	1018	16 Q8FWE0	Q8fwe0 corynebacte
7	7	35.0	1019	16 Q8NM51	Q8nm51 corynebacte
8	7	35.0	1369	10 Q8SAZ5	Q8saz5 oryza sativ
9	7	35.0	1865	10 Q81909	Q81909 arabidopsis
10	6	30.0	32	8 Q36494	Q36494 farfantepen
11	6	30.0	35	3 Q874R5	Q874r5 schizosacch
12	6	30.0	50	16 Q989Z9	Q989z9 rhizobium 1
13	6	30.0	57	9 Q858U2	Q858u2 yersinia pe
14	6	30.0	59	8 Q85X72	Q85x72 pinus korai
15	6	30.0	78	12 Q91R80	Q91r80 hibiscus ch
16	6	30.0	87	16 Q9A719	Q9a719 caulobacter

17	6	30.0	93	2 Q8KIN1	Q8kin1 rhizobium e
18	6	30.0	101	3 Q9P7G3	Q9p7g3 schizosacch
19	6	30.0	108	12 Q8B9M9	Q8b9m9 rachiplusia
20	6	30.0	110	12 Q92387	Q92387 bombyx mori
21	6	30.0	111	17 Q59222	Q59222 pyrococcus
22	6	30.0	112	10 Q9A501	Q9a501 oryza sativ
23	6	30.0	139	5 Q9N4G1	Q9n4g1 caenorhabdi
24	6	30.0	159	16 Q98KG3	Q98kg3 rhizobium 1
25	6	30.0	165	16 Q8XRZ3	Q8xrz3 ralstonia s
26	6	30.0	171	17 Q8TYN9	Q8ty9 methanopyru
27	6	30.0	180	17 Q97V82	Q97v82 sulfolobus
28	6	30.0	182	12 Q80AH4	Q80ah4 hepatitis b
29	6	30.0	183	16 Q9HXU2	Q9hxu2 pseudomonas
30	6	30.0	192	4 Q8N1Z0	Q8n1z0 homo sapien
31	6	30.0	192	10 Q7XR59	Q7xr59 oryza sativ
32	6	30.0	194	4 Q436Z5	Q436z5 homo sapien
33	6	30.0	202	17 Q8ZVF2	Q8zvf2 pyrobaculum
34	6	30.0	203	10 Q9AT41	Q9at41 populus tre
35	6	30.0	205	16 Q8X7K5	Q8x7k5 escherichia
36	6	30.0	214	3 Q9UTW4	Q9utw4 schizosacch
37	6	30.0	216	11 Q8C9S7	Q8c9s7 mus musculu
38	6	30.0	217	16 Q8DTV7	Q8dtv7 streptococc
39	6	30.0	223	16 Q8XTY6	Q8xtv6 ralstonia s
40	6	30.0	226	16 Q05187	Q05187 bacillus su
41	6	30.0	246	16 Q9XDD7	Q9xdd7 brucella me
42	6	30.0	247	17 Q26219	Q26219 methanobact
43	6	30.0	254	10 Q9FG62	Q9fg62 arabidopsis
44	6	30.0	255	4 Q9H3S2	Q9h3s2 homo sapien
45	6	30.0	255	4 Q7Z6E7	Q7z6e7 homo sapien

ALIGNMENTS

RESULT 1

Q9JLY8 ID Q9JLY8 PRELIMINARY; PRT; 355 AA.
AC Q9JLY8; DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Macrophage inflammatory protein-1 alpha receptor.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN NCBI_TaxID=10116;
RN [1]
RC STRAIN=CD.
RX MEDLINE=20555330; PubMed=11091494;
RA Waller A., Nayee P., Czaplowski L.G.;
RT "Identification and characterization of a rat macrophage inflammatory
protein-1 alpha receptor."
RL J. Hematother. Stem Cell Res. 9:703-710(2000).
DR EMBL: AF119381; AAF34340.1; --
DR GO: GO:0016021; C:integral to membrane; IEA.
DR GO: GO:0004872; F:receptor activity; IEA.
DR GO: GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR InterPro: IPR0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR Pfam: PF00001; 7tm.1; 1.
DR PRINTS: PR00237; GPCR_RHODPSN.
DR PROSITE: PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE: PS00262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 355 AA; 40838 MW; 2FEB8661D1B6E075 CRC64;

Query Match 35.0%; Score 7; DB 11; Length 355;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLWTPY 15
|||||

Db 249 FLLWTPY 255

RESULT 2

Q8BVM4 PRELIMINARY; PRT; 355 AA.

AC Q8BVM4; 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Chemokine.

GN CCRI.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Skin;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium.

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RL 60,770 full-length cDNAs.";

RL Nature 420:563-573 (2002).

DR EMBL; AK076275; BAC36284.1; --

DR MGD; MGI:104618; Ccr1.

DR GO; GO:0016493; F-C-C chemokine receptor activity; IDA.

DR GO; GO:0005515; F:protein binding; IPI.

DR GO; GO:0005954; P:immune cell chemotaxis; IDA.

DR GO; GO:0006954; P:inflammatory response; IMP.

DR GO; GO:0030099; P:myeloid blood cell differentiation; IMP.

DR InterPro; IPR000276; GPCR_Rhodpsn.

DR Pfam; PF00001; 7tm1.1; 1.

DR PRINTS; PR00237; GPCRHHODPSN.

DR PROSITE; PS00237; G-PROTEIN RECP Fl_1; 1.

DR PROSITE; PS00262; G-PROTEIN RECP Fl_2; 1.

DR GO; GO:0030099; P:myeloid blood cell differentiation; IMP.

DR InterPro; IPR000276; GPCR_Rhodpsn.

DR Pfam; PF00001; 7tm1.1; 1.

DR PRINTS; PR00237; GPCRHHODPSN.

DR PROSITE; PS00237; G-PROTEIN RECP Fl_1; 1.

DR PROSITE; PS00262; G-PROTEIN RECP Fl_2; 1.

SQ SEQUENCE 355 AA; 40908 MW; F41B2ED5150F7725 CRC64;

Query Match 35.0%; Score 7; DB 11; Length 355;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLLWTPY 15

Db 249 FLLWTPY 255

RESULT 3

Q8BVM4 PRELIMINARY; PRT; 355 AA.

AC Q8BVM4; 23, Created)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)

DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)

DE Chemokine.

GN CCRI.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Forelimb;

RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium.

RA the RIKEN Genome Exploration Research Group Phase I & II Team;

RT "Analysis of the mouse transcriptome based on functional annotation of

RL 60,770 full-length cDNAs.";

RL Nature 420:563-573 (2002).

DR EMBL; AK031109; BAC27256.1; --

DR MGD; MGI:104618; Ccr1.

DR GO; GO:0016493; F-C-C chemokine receptor activity; IDA.

DR GO; GO:0005515; F:protein binding; IPI.

DR GO; GO:0030595; P:immune cell chemotaxis; IDA.

DR GO; GO:0006954; P:inflammatory response; IMP.

DR GO; GO:0030099; P:myeloid blood cell differentiation; IMP.

DR InterPro; IPR000276; GPCR_Rhodpsn.

DR Pfam; PF00001; 7tm1.1; 1.

DR PRINTS; PR00237; GPCRHHODPSN.

DR PROSITE; PS00237; G-PROTEIN RECP Fl_1; 1.

DR PROSITE; PS00262; G-PROTEIN RECP Fl_2; 1.

SQ SEQUENCE 355 AA; 40911 MW; E0B4E87FABC4ED CRC64;

Query Match 35.0%; Score 7; DB 11; Length 355;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 FLLWTPY 15

Db 249 FLLWTPY 255

RESULT 4

Q8TUV5 PRELIMINARY; PRT; 468 AA.

AC Q8TUV5; 21, Created)

DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)

DE Uncharacterized membrane protein.

GN MK1648.

OS Methanopyrus kandleri.

OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;

OC Methanopyrus.

OX NCBI_TaxID=2320;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AV19 / DSM 6324 / JCM 9639;

RX MEDLINE=21927647; PubMed=11930014;

RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,

RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,

RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,

RA Malykh A.G., Koonin E.V., Kozlyavkin S.A.

RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19 and monophyly of archaeal methanogens.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).

DR EMBL; AB010453; AAM02861.1; --

KW Complete proteome.

SQ SEQUENCE 468 AA; 51789 MW; AFC9C3586D4A4E78 CRC64;

Query Match 35.0%; Score 7; DB 17; Length 468;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SVLSFLL 11

Db 303 SVLSFLL 309

RESULT 5

Q74247 PRELIMINARY; PRT; 525 AA.

AC Q74247; 08, Created)

DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)

DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)

DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)

DE Inositol 1-phosphate synthase (EC 5.5.1.4).

GN INOL.

OS Pichia pastoris (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Pichia.

OX NCBI_TaxID=4922;

RN [1]

RP SEQUENCE OF 21-525 FROM N.A.

RX MEDLINE=99392452; PubMed=10463156;


```

Best Local Similarity 100.0%; Pred. NO. 50;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 4 HSVLSFL 10
DB 178 HSVLSFL 184
|||||

RESULT 10
Q36494
ID Q36494 PRELIMINARY; PRT; 32 AA.
AC Q36494;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE NADH dehydrogenase subunit 5 (Fragment).
OS Farfantepenaeus notialis (Southern pink shrimp).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Dendrobranchiata; Penaeoidea;
OC Penaeidae; Farfantepenaeus.
NCBI_TaxID=44108;
RN [1]
RP SEQUENCE FROM N.A.
RA Garcia-Machado E., Dennebouv N., Oliva-Suarez M., Mounolou J.C.,
RT "Partial sequence of the shrimp Penaeus notialis mitochondrial
RL genome.";
RN [2]
RP C. R. Acad. Sci., D, Sci. Nat. 319:473-486(1996).
RA Garcia-Machado E.;
RL Thesis (1997), Paris-Sud, UFR d'Orsay, France.
RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE=99298297; PubMed=10368442;
RL Garcia-Machado E., Pempera M., Dennebouv N., Oliva-Suarez M.,
RT Mounolou J.C., Monnerot M.;
RL "Mitochondrial genes collectively suggest the paraphyly of Crustacea
RL with respect to Insecta";
RL J. Mol. Evol. 49:142-149(1999).
DR EMBL; X84350; CAB40370.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
FT MITOCHONDRION.
FT NON_TER 1
FT NON_TER 32
FT NON_TER 32
SQ SEQUENCE 32 AA; 3783 MW; 32A62974B4937800 CRC64;

Query Match 30.0%; Score 6; DB 8; Length 32;
Best Local Similarity 100.0%; Pred. NO. 18;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 LSFLLW 12
DB 18 LSFLLW 23
|||||

RESULT 11
Q874R5
ID Q874R5 PRELIMINARY; PRT; 35 AA.
AC Q874R5;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Mde5 protein (Fragment).
OS MDE5.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z98944; CAD62442.1; -.
FT NON TER 35
SQ SEQUENCE 35 AA; 4064 MW; 3C800B48D391AC2E CRC64;

Query Match      30.0%; Score 6; DB 3; Length 35;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFL 11
Db 13 VLSFL 18

RESULT 12
Q989Z9 PRELIMINARY; PRT; 50 AA.
AC Q989Z9
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Transposase.
GN MSR6219.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP03008; BAB52545.1; -.
KW Complete proteome.
SQ SEQUENCE 50 AA; 5204 MW; EFD8E5BD1587B0B CRC64;

Query Match      30.0%; Score 6; DB 16; Length 50;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SVLSFL 10
Db 11 SVLSFL 16

RESULT 13
Q858J2 PRELIMINARY; PRT; 57 AA.
AC Q858J2
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Protein 19.3.
OS Yersinia pestis phase ph1A1122.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=227720;
RN [1]
RP SEQUENCE FROM N.A.
RA Garcia E., Elliott J.M., Ramanculov E., Chain P.S., Chu M.C.,
RA Molineux I.J.;
RT "The genome sequence of Yersinia pestis bacteriophage PhiA1122 reveals
RT an intimate history with the coliphage T3.";

```

```

RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY247822; AAP20547.1; -.
SQ SEQUENCE 57 AA; 6418 MW; A0A744FDB0EFB564 CRC64;

Query Match      30.0%; Score 6; DB 9; Length 57;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SVLSFL 10
Db 38 SVLSFL 43

RESULT 14
Q85X72 PRELIMINARY; PRT; 59 AA.
AC Q85X72
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PSII K protein.
GN PSBK.
OS Pinus koraiensis (Korean pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
OX NCBI_TaxID=88728;
RN [1]
RP SEQUENCE FROM N.A.
RA Noh E.W., Lee J.S., Choi Y.I., Han M.S., Yi Y.S., Han S.U.;
RT "Complete nucleotide sequence of Pinus koraiensis.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY228468; AAO73990.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
DR GO; GO:0009538; C:photosystem I reaction center; IEA.
DR GO; GO:0015979; P:photosynthesis; IEA.
DR InterPro; IPR003687; PSII_PSBK.
DR Pfam; PF02533; PsbK; 1.
KW Chloroplast.
SQ SEQUENCE 59 AA; 6672 MW; 5AB7C7F6D6701FED CRC64;

Query Match      30.0%; Score 6; DB 8; Length 59;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 VLSFL 11
Db 43 VLSFL 48

RESULT 15
Q9IR80 PRELIMINARY; PRT; 78 AA.
AC Q9IR80
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Putative movement protein p9.
OS Hibiscus chlorotic ringspot virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;
OC Carmovirus.
OX NCBI_TaxID=53181;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang M., Koh D.C.Y., Weng L., Chang M.L., Yap Y.K., Zhang Y.K.,
RA Wong S.M.;
RT "Complete nucleotide sequence and genome organization of hibiscus
RT chlorotic ringspot virus, a new member of the genus carmovirus:
RT evidence for the presence and expression of two novel open reading
RT frames.";
RL J. Virol. 74:3149-3155(2000).
DR EMBL; X86448; CAB81770.1; -.

```

SQ SEQUENCE 78 AA; 8995 MW; 1E18EB61BB2C5CAB CRC64;
Query Match 30.0%; Score 6; DB 12; Length 78;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 VLSFLL 11
|||
Db 7 VLSFLL 12

Search completed: July 19, 2004, 17:14:20
Job time : 39 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 21, 2004, 00:04:45 ; Search time 1705 Seconds
(without alignments)
508.423 Million cell updates/sec

Title: US-10-799-747-116
Perfect score: 99
Sequence: 1 MAHSVLSFLWTPYALKSX 20

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPTO spool_p/US10799747/runat 19072004 161416 21440/app query.fasta_1.199
-DB=GenEmbl -QFMT=fastap -SUPFIX=rg -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10799747 @CGN 1.1 5600 @runat 19072004 161416 21440 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_ox.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length	%			
1	98	99.0	1434	6	BD078422	BD078422	101 human
2	98	99.0	3452	9	BC033650	BC033650	Homo sapi
3	98	99.0	3522	9	BSM804706	BSM804706	AL833393 Homo sapi
C 4	98	99.0	175081	9	AC021151	AC021151	AC021151 Homo sapi
C 5	98	99.0	281662	2	AC117374	AC117374	AC117374 Homo sapi
C 6	90	90.9	2020	9	HSB800227	HSB800227	AL049442 Homo sapi
C 7	58	58.6	194420	2	AC124133	AC124133	AC124133 Mus muscu
C 8	58	58.6	244669	2	AC096099	AC096099	AC096099 Rattus no
C 9	58	58.6	245830	2	AC140313	AC140313	AC140313 Mus muscu
C 10	57	57.6	156471	2	AC013568	AC013568	AC013568 Homo sapi
C 11	57	57.6	192867	9	AL358815	AL358815	AL358815 Human DNA
C 12	56	56.6	81715	9	AC024680	AC024680	AC024680 Homo sapi
C 13	56	56.6	165127	9	AC091047	AC091047	AC091047 Homo sapi
C 14	56	56.6	233717	2	AC140381	AC140381	AC140381 Mus muscu
C 15	55	55.6	98558	9	AL137138	AL137138	AL137138 Human DNA
16	54.5	55.1	138187	10	MMMH461	MMMH461	AF027865 Mus muscu
17	54	54.5	51559	2	AC099861	AC099861	AC099861 Mus muscu
18	54	54.5	137930	2	AC141193	AC141193	AC141193 Rattus no
19	54	54.5	201465	2	AC122424	AC122424	AC122424 Mus muscu
C 20	54	54.5	226253	2	AC133434	AC133434	AC133434 Rattus no
C 21	54	54.5	227323	2	AC117350	AC117350	AC117350 Rattus no
C 22	54	54.5	272007	2	AC110094	AC110094	AC110094 Rattus no
C 23	53	53.5	30040	2	AC015491	AC015491	AC015491 Homo sapi
C 24	53	53.5	60464	2	AL359673	AL359673	AL359673 Homo sapi
C 25	53	53.5	60828	9	HS1191B2	HS1191B2	AL022237 Human DNA
C 26	53	53.5	63007	2	AC037446	AC037446	AC037446 Homo sapi
C 27	53	53.5	63731	2	AL360016	AL360016	Continuation (5 of
C 28	53	53.5	115198	2	AL139223	AL139223	Continuation (5 of
C 29	53	53.5	135571	2	AC116710	AC116710	AC116710 Homo sapi
C 30	53	53.5	145111	2	AL390245	AL390245	AL390245 Homo sapi
C 31	53	53.5	148679	9	AC099788	AC099788	AC099788 Homo sapi
C 32	53	53.5	148729	9	AC119724	AC119724	AC119724 Homo sapi
C 33	53	53.5	153615	2	AC111043	AC111043	AC111043 Mus muscu
C 34	53	53.5	176308	2	AC124005	AC124005	AC124005 Mus muscu
C 35	53	53.5	176684	2	AC023127	AC023127	AC023127 Homo sapi
C 36	53	53.5	204495	10	AL591884	AL591884	AL591884 Mouse DNA
C 37	53	53.5	209513	2	AC122952	AC122952	AC122952 Rattus no
C 38	53	53.5	227363	2	AL592002	AL592002	AL592002 Mus muscu
C 39	53	53.5	277861	2	HSAC000406	HSAC000406	AC000406 Homo sapi
C 40	52	52.5	4194	9	AK126784	AK126784	AK126784 Homo sapi
C 41	52	52.5	49743	2	AC100574	AC100574	AC100574 Mus muscu
C 42	52	52.5	61993	2	AC137891	AC137891	AC137891 Homo sapi
C 43	52	52.5	62550	2	AC136338	AC136338	AC136338 Homo sapi
C 44	52	52.5	66251	2	AC115984	AC115984	AC115984 Homo sapi
45	52	52.5	94308	9	AC005480	AC005480	AC005480 Homo sapi

ALIGNMENTS

RESULT 1

```

BD078422
LOCUS       BD078422               1434 bp    DNA        linear    PAT 27-AUG-2002
DEFINITION  101 human secretory proteins.
ACCESSION   BD078422
VERSION     BD078422.1 GI:22624025
KEYWORDS    JP 2001519156-A/11
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1434)
AUTHORS    Duan,R.D., Florence,K.A., Rosen,C.A., Ruben,S.M., Greene,J.M.,
            Young,P., Ferrie,A.M., Yu,G.L., Janat,P., Ni,J., Carter,K.C.,
            Endress,G.A., Feng,P., Lafleur,D.W. and Shi,Y.
            101 human secretory proteins
            Patent: JP 2001519156-A 11 23-OCT-2001;
            HUMAN GENOME SCIENCES INC
COMMENT     OS Homo sapiens (human)
            PN JP 2001519156-A/11
            PD 23-OCT-2001
            PF 01-OCT-1998 JP 2000515006
            PR 02-OCT-1997 US 60/060837,02-OCT-1997 US 60/060862 PR
            02-OCT-1997 US 60/060839,02-OCT-1997 US 60/060866 PR
            02-OCT-1997 US 60/060843,02-OCT-1997 US 60/060836 PR
            02-OCT-1997 US 60/060838,02-OCT-1997 US 60/060874 PR
            02-OCT-1997 US 60/060833,02-OCT-1997 US 60/060884 PR
            02-OCT-1997 US 60/060880
            PI ROXANNE D DUAN,KIMBERLY A FLORENCE,CRAIG A ROSEN,STEVEN M PI
            RUBEN,
            PI JOHN M GREENE,PAUL YOUNG,ANN M FERRIE,GUO
            LIANG YU,FOUAD JANAT,
            PI JIAN NI,
            PI KENNETH C CARTER,GREGORY A ENDRESS,PING FENG,DAVID W LALEUR,
            PI YANGGU SHI
            PC C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P25/00,A61P35/00,
            A61P37/00,
            PC A61P43/00,C07K14/47,C07K16/00,C12N5/10,C12P21/02,C12Q1/68, PC
            G01N33/53,
            PC G01N33/53,G01N33/566,C12N15/00,A61K37/02,C12N5/00 CC 101
            human secretory proteins
            FH Key Location/Qualifiers
            FT source 1. 1434
            /organism="Homo sapiens (human)".

FEATURES             source
source              1. 1434
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:         7.78e-08      Length:      1434
Score:             98.00         Matches:     19
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%   Mismatches:  0
Query Match:       98.99%        Indels:      0
DB:                6            Gaps:        0

US-10-799-747-116 (1-20) x BD078422 (1-1434)

QY      1 MetAlaAlaHisSerValLeuSerPheLeuLeuTriPThrProTyraLaLeuLysSer 19
Db      507 AUGGACGCCATTTCAGTCTTGAGTTTCTCTCTGGACACCTATGCTCTGAATCA 563

RESULT 2
BC033650
LOCUS       BC033650               3452 bp    mRNA        linear    PRI 06-OCT-2003
DEFINITION  Homo sapiens hypothetical protein FLJ14431, mRNA (cdna clone
            MGC:44889 IMAGE:5574637), complete cds.
ACCESSION   BC033650
VERSION     BC033650.1 GI:21707066
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)

```

ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE

PUBMED

REFERENCE
AUTHORSTITLE
JOURNALREMARK
COMMENTFEATURES
source

gene

CDS

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3452)

Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,

Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,

Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,

Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,

Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,

Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,

Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,

Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,

McKernan,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,

Abramson,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,

Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,

Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,

Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,

Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,

Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,

Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,

Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smaluse,D.E.,

Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.

Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

2238257

12477932

2 (bases 1 to 3452)

Strausberg,R.

Direct Submission

Submitted (02-JUL-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs@mail.nih.gov

Tissue procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),

Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>Contact: nisc.mgc@nih.gov

Akter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,

Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,

Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,

Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,

Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,

McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,

Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,

Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LIML at: <http://image.llnl.gov>

Series: IRAK Plate: 69 Row: n Column: 19

This clone was selected for full length sequencing because it

passed the following selection criteria: matched mRNA gi: 14249445.

Location/Qualifiers

1. 3452

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:44889 IMAGE:5574637"

/tissue_type="Duodenum, adenocarcinoma"

/clone_lib="NIH MGC 88"

/lab_host="DH10B"

/note="Vector: pCMV-Sport6"

1. 3452

/gene="FLJ14431"

/db_xref="LocusID:84869"

159. 872


```

/codon_start=1
/product="FLJ14431 protein"
/protein_id="AAH33650.1"
/db_xref="GI:21707067"
/db_xref="LocusID:84869"
/translation="MDKVCAGGSGRGIGRAVAQLMARKGYRLAVIARNLEGAKAAG
DIGGDHLASCDVAKHDVQNTFELEKHLGRVFNFAAGINRDLGSLVKTEDMVS
QHTNLGSLMTCIAAMRTMIQQQSGIVNGSVGLKNGSQSVYSASKGLGVFSR
ALAKEVARKIRVNVVAPGVHTDWTDKLKEELKKNIPLGRFGETIEVAHAVVFLE
SPYITGHVLVDVGGQLIL"
misc_feature
159..857
/note="adh_short; Region: short chain dehydrogenase. This
family contains a wide variety of dehydrogenases"
/db_xref="CDD:pfam00106"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 2,11e-07 Length: 3452
Score: 98.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.99% Indels: 0
DB: 9 Gaps: 0

```

US-10-799-747-116 (1-20) x BC033650 (1-3452)

```

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTriThrProTyrrAlaLeuLySer 19
|||||
Db 2542 ATGGCAGCCCAATTCAGTCTTGAGTTTCTTCTCGACACCTTATGCTCTGAAATCA 2598
|||||

RESULT 3
HSM804706
LOCUS Homo sapiens mRNA; cDNA DKFp762K109 (from clone DKFp762K109).
ACCESSION AL833393
VERSION AL833393.1 GI:21734029
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3522)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann,S.
JOURNAL Direct Submission
TITLE Submitted (09-JUL-2002) 1, D-85764 Neuberberg, GERMANY
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Medicinix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. This clone
(DKFp762K109) is available at the RZPD in Berlin. Please contact
the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
Location/Qualifiers
1..3522
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFp762K109"
/db_xref="taxon:9606"
/clone="DKFp762K109"
/issue_type="melanoma (MeWo cell line)"
/clone_lib="762 (synonym: hmel2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
3420..3425
polya_signal
3420..3425
polya_site
3440

```

ORIGIN

```

Alignment Scores:
Pred. No.: 2,16e-07 Length: 3522
Score: 98.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0

```

```

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.99% Indels: 0
DB: 9 Gaps: 0

```

US-10-799-747-116 (1-20) x HSM804706 (1-3522)

```

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTriThrProTyrrAlaLeuLySer 19
|||||
Db 2552 ATGGCAGCCCAATTCAGTCTTGAGTTTCTTCTCGACACCTTATGCTCTGAAATCA 2608
|||||

```

RESULT 4

AC021151/c

LOCUS Homo sapiens BAC clone RP11-483A20 from 4, complete sequence.

DEFINITION AC021151

ACCESSION AC021151

VERSION AC021151.8 GI:15145598

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 175081)

AUTHORS Sulston,J.E. and Waterston,R.

TITLE Toward a complete human genome sequence

JOURNAL Genome Res. 8 (11), 1097-1108 (1998)

MEDLINE 99063792

PUBMED 9847074

REFERENCE 2 (bases 1 to 175081)

AUTHORS Kyung,K. and Abbott,A.

TITLE The sequence of Homo sapiens BAC clone RP11-483A20

JOURNAL Unpublished (2001)

REFERENCE 3 (bases 1 to 175081)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

REFERENCE 4 (bases 1 to 175081)

AUTHORS Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (09-AUG-2001) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

REFERENCE 5 (bases 1 to 175081)

AUTHORS Waterston,R.

TITLE Direct Submission

JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Aug 9, 2001 this sequence version replaced gi:13877272.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: http://genome.wustl.edu/gsc

Contact: saplense@wustl.edu

----- Summary Statistics

Center project name: H_NH0483A20

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D.

McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Onoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenno, M., Cataneese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-36G9. Actual start of this clone is at base position 1 of RP11-483A20; actual end is at base position 175081 of RP11-483A20.

FEATURES

source

```
1. .175081
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="4"
   /map="4"
   /clone="RP11-483A20"
   /clone_lib="RPCT-11"
```

repeat_region

```
1. .3267
   /rpt_family="L1"
3270. .3453
   /rpt_family="Alu"
3434. .3464
   /rpt_family="(A)n"
3455. .4161
   /rpt_family="L1"
```

misc_feature

```
7949. .7956
   /notes="match to EST AW006057 (NID:g5854835) wz81d04.x1"
```

repeat_region

```
10197. .10489
   /rpt_family="L1"
10480. .10501
   /rpt_family="AT_rich"
10495. .10665
   /rpt_family="L1"
10731. .11051
   /rpt_family="L1"
11052. .11576
   /rpt_family="L1"
```

repeat_region

```
11564. .11598
   /rpt_family="(A)n"
11642. .12324
   /rpt_family="L1"
11769. .11794
   /rpt_family="AT_rich"
```

repeat_region

```
11939. .11962
   /rpt_family="AT_rich"
12335. .12412
   /rpt_family="L1"
12392. .12621
   /rpt_family="L1"
```

repeat_region

```
12788. .13096
   /rpt_family="Alu"
13069. .13102
   /rpt_family="(CAAAA)n"
13325. .13667
   /rpt_family="Alu"
```

repeat_region

```
13444. .13484
   /rpt_family="(TAAA)n"
13640. .13686
   /rpt_family="(GAAAA)n"
14249. .14422
   /rpt_family="MERL_type"
```

repeat_region

```
14980. .15182
   /rpt_family="L1"
```

```
repeat_region 15260. .15492
               /rpt_family="L1"
repeat_region 15483. .15985
               /rpt_family="MERL_type"
repeat_region 15986. .16105
               /rpt_family="L1"
repeat_region 16117. .16251
               /rpt_family="L1"
repeat_region 16288. .16356
               /rpt_family="L1"
repeat_region 16373. .16545
               /rpt_family="L1"
repeat_region 16550. .18773
               /rpt_family="L1"
repeat_region 18770. .20218
               /rpt_family="L1"
repeat_region 19228. .19346
               /rpt_family="T-rich"
repeat_region 20219. .20514
               /rpt_family="Alu"
repeat_region 20515. .22484
               /rpt_family="L1"
repeat_region 22584. .22882
               /rpt_family="Alu"
repeat_region 22748. .22768
               /rpt_family="AT_rich"
repeat_region 22883. .23460
               /rpt_family="L1"
repeat_region 23461. .23747
               /rpt_family="MaLR"
repeat_region 23749. .24514
               /rpt_family="L1"
misc_feature 24456
               /notes="match to EST AW902102 (NID:g8066307)"
misc_feature 24488. .24797
               /notes="match to EST AW902102 (NID:g8066307)"
repeat_region 24576. .24625
               /rpt_family="Mariner"
repeat_region 25484. .25787
               /rpt_family="Alu"
repeat_region 25758. .25787
               /rpt_family="(A)n"
repeat_region 25796. .26014
               /rpt_family="MIR"
repeat_region 26657. .26889
               /rpt_family="L1"
repeat_region 26890. .27205
               /rpt_family="MaLR"
repeat_region 27206. .27439
               /rpt_family="Alu"
repeat_region 27440. .27490
               /rpt_family="MaLR"
repeat_region 27491. .28717
               /rpt_family="L1"
repeat_region 28718. .28742
               /rpt_family="(T)n"
misc_feature 28735. .28771
               /notes="match to EST AA954400 (NID:g3118095) 0001f01.s1"
misc_feature 28747. .28826
               /notes="match to EST AW006057 (NID:g5854835) wz81d04.x1"
repeat_region 29022. .29308
               /rpt_family="L1"
misc_feature 29384. .30105
               /notes="match to EST BF680971 (NID:gl1954866)"
misc_feature 29687. .30351
               /notes="match to EST BG429938 (NID:gl13336444)"
```

Alignment Scores:

Pred. No.:	1.81e-05	Length:	175081
Score:	98.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.99%	Indels:	0

DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x AC021151 (1-175081)

QY 1 MetAlaAlaHisSerValLeuSerPheLeuLeuTrrPheProTyrAlaLeuLysSer 19
|||||

Db 57641 ATGGCAGCCATTCAGTCTTGAGTTTCTCTCTGACACCTTATGCTTGAAATCA 57585

RESULT 5
AC117374/c
LOCUS
DEFINITION Homo sapiens chromosome 12 clone RP11-19D19, *** SEQUENCING IN
AC117374 AC011694
VERSION AC117374.1 GI:20127352
KEYWORDS HTG; HTGS PHASE1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 281662)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Binage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gortell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Hollaway,C., Hollins,B.,
Homi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.B.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louiseged,H.,
Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,B., Mawhiney,B., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vazquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 281662)
Worley,K.C.
Direct Submission
Submitted (10-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 281662)
Worley,K.C.
Direct Submission
Submitted (02-MAY-2002) Human Genome Sequencing Center, Department

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 28, 2002 this sequence version replaced gi:10045385.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HCN2
Center clone name: RP11-19D19
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 273063 bases at least Q40
Consensus quality: 294336 bases at least Q30
Consensus quality: 306942 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2464: contig of 2464 bp in length
* 2465 2564: gap of unknown length
* 2565 4603: contig of 2039 bp in length
* 4604 4703: gap of unknown length
* 4704 7113: contig of 2410 bp in length
* 7114 7213: gap of unknown length
* 7214 9517: contig of 2304 bp in length
* 9518 9617: gap of unknown length
* 9618 12107: contig of 2490 bp in length
* 12108 12207: gap of unknown length
* 12208 15014: contig of 2807 bp in length
* 15015 15114: gap of unknown length
* 15115 17331: contig of 2717 bp in length
* 17332 17931: gap of unknown length
* 17932 20480: contig of 2549 bp in length
* 20481 20580: gap of unknown length
* 20581 22700: contig of 2120 bp in length
* 22701 22800: gap of unknown length
* 22801 24909: contig of 2109 bp in length
* 24910 25009: gap of unknown length
* 25010 27313: contig of 2304 bp in length
* 27314 27413: gap of unknown length
* 27414 29798: contig of 2385 bp in length
* 29799 29898: gap of unknown length
* 29899 33466: contig of 3568 bp in length
* 33467 33666: gap of unknown length
* 33667 36779: contig of 3113 bp in length
* 36780 40472: contig of 3693 bp in length
* 40473 40572: gap of unknown length
* 40573 42683: contig of 2111 bp in length
* 42684 42783: gap of unknown length
* 42784 46319: contig of 3536 bp in length
* 46320 46419: gap of unknown length
* 46420 49507: contig of 3088 bp in length
* 49508 49607: gap of unknown length
* 49608 53212: contig of 3605 bp in length
* 53213 53312: gap of unknown length
* 53313 57267: contig of 3955 bp in length
* 57268 57367: gap of unknown length
* 57368 60444: contig of 3077 bp in length
* 60445 60544: gap of unknown length
* 60545 65859: contig of 5315 bp in length
* 65860 65959: gap of unknown length

```

* 65960 69038: contig of 3079 bp in length
* 69039 69138: gap of unknown length
* 69139 73445: contig of 4307 bp in length
* 73446 73545: gap of unknown length
* 73546 78107: contig of 4562 bp in length
* 78108 82504: gap of unknown length
* 82505 82604: contig of 4297 bp in length
* 82605 86976: gap of unknown length
* 86977 87076: contig of 4372 bp in length
* 87077 91315: contig of 4239 bp in length
* 91316 91415: gap of unknown length
* 91416 95084: contig of 3669 bp in length
* 95085 95184: gap of unknown length
* 95185 102348: contig of 7164 bp in length
* 102349 102448: gap of unknown length
* 102449 105973: contig of 3525 bp in length
* 105974 106073: gap of unknown length
* 106074 112891: contig of 6818 bp in length
* 112892 112991: gap of unknown length
* 112992 116221: contig of 3230 bp in length
* 116222 116321: gap of unknown length
* 116322 119798: contig of 3477 bp in length
* 119799 119898: gap of unknown length
* 119900 124870: contig of 4972 bp in length
* 124871 124970: gap of unknown length
* 124971 130790: contig of 5820 bp in length
* 130791 130890: gap of unknown length
* 130891 134639: contig of 3749 bp in length
* 134640 134739: gap of unknown length
* 134740 140637: contig of 5898 bp in length
* 140638 140737: gap of unknown length
* 140738 145541: contig of 4804 bp in length
* 145542 145641: gap of unknown length
* 145642 155057: contig of 9416 bp in length
* 155058 155157: gap of unknown length
* 155158 164357: contig of 9200 bp in length
* 164358 164457: gap of unknown length
* 164458 174832: contig of 10435 bp in length
* 174833 174932: gap of unknown length
* 174933 184400: contig of 9408 bp in length
* 184401 184500: gap of unknown length
* 184501 198385: contig of 13885 bp in length
* 198386 198485: gap of unknown length
* 198486 216436: contig of 17951 bp in length
* 216437 216537: gap of unknown length
* 216538 22916: contig of 16380 bp in length
* 22917 233016: gap of unknown length
* 233017 255437: contig of 22421 bp in length
* 255438 255537: gap of unknown length
* 255538 281662: contig of 26125 bp in length.

```

FEATURES

source

```

1. .281662
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-19D19"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 3.1e-05 Length: 281662
Score: 98.00 Matches: 19
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 98.99% Indels: 0
DB: 2 Gaps: 0

```

US-10-799-747-116 (1-20) x AC117374 (1-281662)

Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuYsSer 19

Db 145718 ATGGCAGCCCAATTCAGCTTGGATTTCTTCTCTGGACCATATGCTCTGAAATCA 145662

RESULT 6

HSM800227

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

HSM800227 2020 bp mRNA linear PRI 18-FEB-2000
Homo sapiens mRNA; cDNA DKFZp586N1720 (from clone DKFZp586N1720).

AL049442
AL049442.1 GI:4500222

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2020)

Wambutt,R., Heubner,D., Mewes,H.W., Gassenhuber,J. and Wiemann,S.

Direct Submission

Submitted (10-MAR-1999) MIPS, Am Klopferspitze 18a, D-82152

Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;

sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing

consortium of the German Genome Project.

This clone (DKFZp586N1720) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further

information about the clone and the sequencing project is available

at http://www.mips.biochem.mpg.de/proj/cDNA/.

Location/Qualifiers

1. .2020

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="RZPD:DKFZp586N1720"

/db_xref="taxon:9606"

/clone="DKFZp586N1720"

/tissue_type="uterus"

/clone_lib="586 (synonym: hutel). Vector pSport1; host

DH10B; sites NotI + SalI/MluI"

/dev_stage="adult"

1983..1988

polyA_signal

polyA_site

2003

ORIGIN

Alignment Scores:

Pred. No.: 3.26e-06 Length: 2020

Score: 90.00 Matches: 18

Percent Similarity: 94.74% Conservative: 0

Best Local Similarity: 94.74% Mismatches: 1

Query Match: 90.91% Indels: 0

DB: 9 Gaps: 0

US-10-799-747-116 (1-20) x HSM800227 (1-2020)

Qy 1 MetAlaAlaHisSerValLeuSerPheLeuLeuThrProTyrAlaLeuYsSer 19

Db 1115 ATGGCAGCCCAATTCAGCTTGGATTTCTTCTCTGGACCATATGCTCTGAAATCA 1171

RESULT 7

AC124133/c

LOCUS

DEFINITION

IN PROGRESS ***, 2 ordered pieces.

AC124133

AC124133.10 GI:40018787

HTG; HTGS PHASE2; HTGS FULLTOP; HTGS_ACTIVEFIN.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 194420)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus chromosome 15, clone RP23-218L23

Unpublished

2 (bases 1 to 194420)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,